Designing an Eyes-Reduced Document Skimming App for Situational Impairments

by

Taslim Arefin Khan

B. Sc., Bangladesh University of Engineering and Technology, 2015
M. Sc., Bangladesh University of Engineering and Technology, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2019

© Taslim Arefin Khan, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis titled:

Designing an Eyes-Reduced Document Skimming App for Situational Impairments

submitted by **Taslim Arefin Khan** in partial fulfillment of the requirements for the degree of **Master of Science** in **Computer Science**.

Examining Committee:
Joanna McGrenere, Computer Science
Supervisor

Dongwook Yoon, Computer Science
Co-supervisor and Supervisory Committee Member

Abstract

Listening to text using read-out-loud applications is a popular way for people to consume content when their visual attention is situationally impaired (e.g., commuting, walking, tired eyes). However, due to the linear nature of audio, such apps do not support skimming–a non-linear, rapid form of reading–essential for quickly grasping the gist and organization of difficult texts, like academic or professional documents. To support auditory skimming for situational impairment, we: (1) identified the user needs and challenges in auditory skimming through a formative study (N=20), (2) derived the concept of "eyes-reduced" skimming that blends auditory and visual modes of reading, inspired by how participants mixed visual and non-visual interactions, (3) generated a set of design guidelines for eyes-reduced skimming, and (4) designed and evaluated a novel audio skimming app (Skimmer) that embodies the guidelines. Our in-situ preliminary observation study (N=6) suggests that participants are positive about our design and are able to auditorily skim documents. We discuss design implications for eyes-reduced reading, read-out-loud apps, and text-to-speech engines.

Lay Summary

When bleary-eyed or in a bumpy bus, reading text can be difficult or even impossible. People in such *situationally impaired* conditions often listen to a text-to-speech (TTS) app that narrates the given text out loud. Listening to a machine-generated narration from start to end is a good way for understanding easy content, such as fiction. In contrast, when people read more difficult content such as research papers with their eyes, they tend to "skim read" by jumping between pages and quickly glancing at selective sections. Existing TTS apps don't fully support audio-based skimming.

To support skim reading during situational impairments, we built Skimmer, an app that helps the user to predominantly listen, but also enables looking at content (e.g., images) when necessary (hence "eyes-reduced" interaction). We evaluated Skimmer in the context of riding a bus and found that users were able to skim documents in an eyes-reduced manner.

Preface

The experiments described in this thesis were conducted with the approval of the UBC Behavioral Research Ethics Board (certifiate number H18-00664).

A significant portion of this thesis has been submitted as a manuscript and is under review in a top-tier conference while submitting this thesis. I am the lead author in that manuscript.

Dr. Joanna McGrenere and Dr. Dongwook Yoon helped ideating and framing the research since its inception in January 2018. Both of them provided supervisory assistance during the time of this research. They provided feedback on the design and development of the research. They also assisted me with the writing of the submitted manuscript mentioned above.

Paul Bucci read this manuscript and provided feedback on writing.

The eDAPT group led by Dr. Joanna McGrenere contributed with feedback and early edits for parts of this manuscript. Several members of the Multimodal User Experience (MUX) lab contributed in this research by participating as pilot participants in the studies, brainstorming ideas, and lending a hand in troubleshooting programming bugs. Ashish Chopra and Mohi Reza helped me address technical problems while implementing Skimmer. David Marino, Godfrey Cheng, Matthew Chun, and Laura Cang helped me capture images and videos during the study.

Table of Contents

Abstract . . . . **iii**

Lay Summary . . . . **iv**

Preface . . . . **v**

Table of Contents . . . . **vi**

List of Tables . . . . **ix**

List of Figures . . . . **x**

Acknowledgements . . . . **xi**

1 Introduction **1**
1.1 Research Questions and Approach 2
1.2 Contributions 6
1.3 Overview 6

2 Related Work **7**
2.1 Situational Impairment and Eyes-free Interaction 7
2.2 Human Listening and Audio Interfaces 9
2.3 Studies on Skimming 10
2.3.1 Document Skimming 10
2.3.2 Audio Skimming 11

3 Understanding User Needs for Auditory Skimming 16
3.1 Method 16
3.1.1 Participants 16
3.1.2 Task 17
3.1.3 Design of the Study 17
3.1.4 Conditions 17
3.1.5 Apparatus 18
3.1.6 Procedure 19
3.1.7 Quantitative Measures 20
3.1.8 Data Collection and Data Analysis 21
3.2 Findings 21

4 Design Guidelines for Eyes-Reduced Skimming 26

5 Design of Skimmer 30
5.1 Eyes-Reduced Skimming Features of Skimmer 30
5.1.1 Basic Navigation for Eyes-Reduced Skimming 30
5.1.2 Designing for Easy to Understand Audio Narration 31
5.1.3 Navigating Between Discourse Markers 31
5.1.4 Understanding the High-level Organization 33
5.1.5 Opt-in to See Visual Content 33
5.1.6 Auditory Cues in Support of Skimming 34
5.1.7 Accommodating Individual Differences in Listening Rates 34
5.2 Design Process 36
5.2.1 Conceptual Design 36
5.2.2 Design and Prototyping Method Refinements 39
5.2.3 Building a Usable Prototype 42

6 Preliminary Evaluation of Skimmer 44
6.1 Method 44
6.2 Findings 45

7 Discussion **49**

8 Limitations **54**

9 Conclusion **55**

Bibliography **56**

A Study Materials Used in Need Finding Study **68**

A.1 Call For Participation 68

A.2 Consent Form for Participants 70

A.3 Study Protocol and Participant Training 74

A.4 Questionnaires and Interview Questions 86

B Study Materials Used in Preliminary Evaluation Study **92**

B.1 Call For Participation 92

B.2 Consent Form for Participants 94

B.3 Study Protocol and Participant Training 98

B.4 Questionnaires and Interview Questions 100

List of Tables

3.1 Articles selected for skimming task in need finding study. 18

List of Figures

1.1 Basic user interface of Skimmer. 4
1.2 User interfaces of Skimmer showing how Skimmer handles numbers and figures/tables in the text. 5

2.1 (Part 1) A comparison of the level of support for various skimming features in different read-out-loud apps. 13
2.2 (Part 2) A comparison of the level of support for various skimming features in different read-out-loud apps. 14
2.3 (Part 3) A comparison of the level of support for various skimming features in different read-out-loud apps. 15

3.1 Participant standing in front of a big screen where we projected ambient bus scenario. 19

5.1 Skimmer's discourse marker navigation. 32
5.2 Low-fidelity paper prototype of Skimmer with different paper screens. 35
5.3 Mapping standard gestures to features. 38
5.4 Refined paper prototype which is evaluated using a wizard-of-oz technique. 40

6.1 Field study to evaluate Skimmer. 45

Acknowledgements

I thank my supervisors - Dr. Joanna McGrenere and Dr. Dongwook Yoon, for their encouragements, patience, and enthusiasm in my research. They have been very supportive of me - often asking and taking good care of my emotional, mental, and financial troubles. I also thank them for deciding to co-supervise me and fund my research all along.

I thank my parents for their endless love and encouragement throughout my Masters. They have always kept me in their prayers and always believed in me. I thank my brother and my sister for also keeping me in their prayers.

I thank my beautiful wife, Saleema Najenin Siddiqua for all the supports, love, encouragement, persistence, and the surprise gifts that made me smile after long hectic days.

I thank all the members from eDAPT group - Shareen Mahmud, Francesco Vitale, Yelim Kim, Izabelle Janzen, Jessalyn Alvina, and Hayley Guillou for their feedback and support all through my Masters. I also thank Paul Bucci, Laura Cang, Soheil Kianzad, Matthew Chun, and Hanieh Shakeri for providing feedback on my research from time to time.

Finally, I thank Sheikh Tanvir for taking good care of me as a roommate and for taking me to some of the most beautiful places in British Columbia and Alberta. I also thank Asma Enayet for occasionally hosting delicious dinners, sometimes at my request, so that I get enough variations in my diet.

All these people kept me moving while miles away from home in Vancouver.

Chapter 1

Introduction

People are consuming media auditorily more than ever before. This trend started with media such as podcasts–content specifically designed for auditory consumption that is typically listened to linearly and is based on pre-recorded human speech. As the quality of text-to-speech (TTS) technology has improved, so has the popularity of listening to machine-generated narrations of textual content using read-out-loud apps (e.g., Mozilla Scout [1], Pocket [2], and Google Go). Using such apps is a good way to consume relatively straight-forward content (e.g., fiction and news articles) that has a linear narrative and can be understood by listening to it from the beginning to the end. However, reading structured, more complex documents often involves a different, particular type of reading—skimming—that the existing read-out-loud apps fail to support.

Skimming, in general, is a rapid, selective, and non-sequential form of reading [3, 4] that is not straightforward to support through the auditory modality which is, in contrast, transient, sequential, and temporal [5]. Skimming requires the reader's eyes to wander over a page [6]. It is also a goal-based reading strategy. People skim to learn the gist of a document, look for specific information in the document, and learn the structural organisation of the document. It involves rapid eye movement over the text, spotting keywords, selectively reading sentences relevant to the goal, and jumping around the text, until the information need is met.

To date, skimming has not been well supported by existing voice reading apps that use TTS.

Yet, skimming is the new norm for reading [7]. The amount of online content alone increases rapidly each year [8], making it impossible for anyone to keep up, even with a small fraction of what is available. In several readerships including business executives, lawyers, doctors, etc., skimming is an important and frequent style of reading for consuming structured texts. Academics, just as one example class of reader, have noted the challenge of keeping up with the literature, given the explosion of publications. Thus, it is not surprising that people are skimming more than ever before, often to assess whether content warrants the effort of a full read [9]. A key restriction is that skimming today requires a person's full visual attention which limits when and where it can be done. Thus, it is not possible to skim in "on-the-go" situations, in the way that one can listen to podcasts or even audiobooks, despite that there is strong evidence calling for better support for "on-the-go" reading [10, 4].

1.1 Research Questions and Approach

In this thesis, our research question is whether it is possible to design a system to effectively support eyes-reduced skim reading, namely a system that combines auditory and visual reading modes. It enables content, such as web pages to academic papers, to be auditorily skimmable eyes-free, but also provides useful visual information should a user choose to selectively glance down at the screen to look at content, such as an image in the text. We use the term eyes-reduced to juxtapose with eyes-free, which has been explored in the literature [11]. Skimming auditorily should help individuals who are situationally impaired, such as those who are on the go, using their visual attention for other things, such as walking or commuting by bus/train but experience motion sickness while reading in a moving vehicle. The commuting context is particularly motivating as commute times are increasing substantially in many parts of the world [12]. At a broader level, people with vision impairments should also get benefit from auditory skimming.

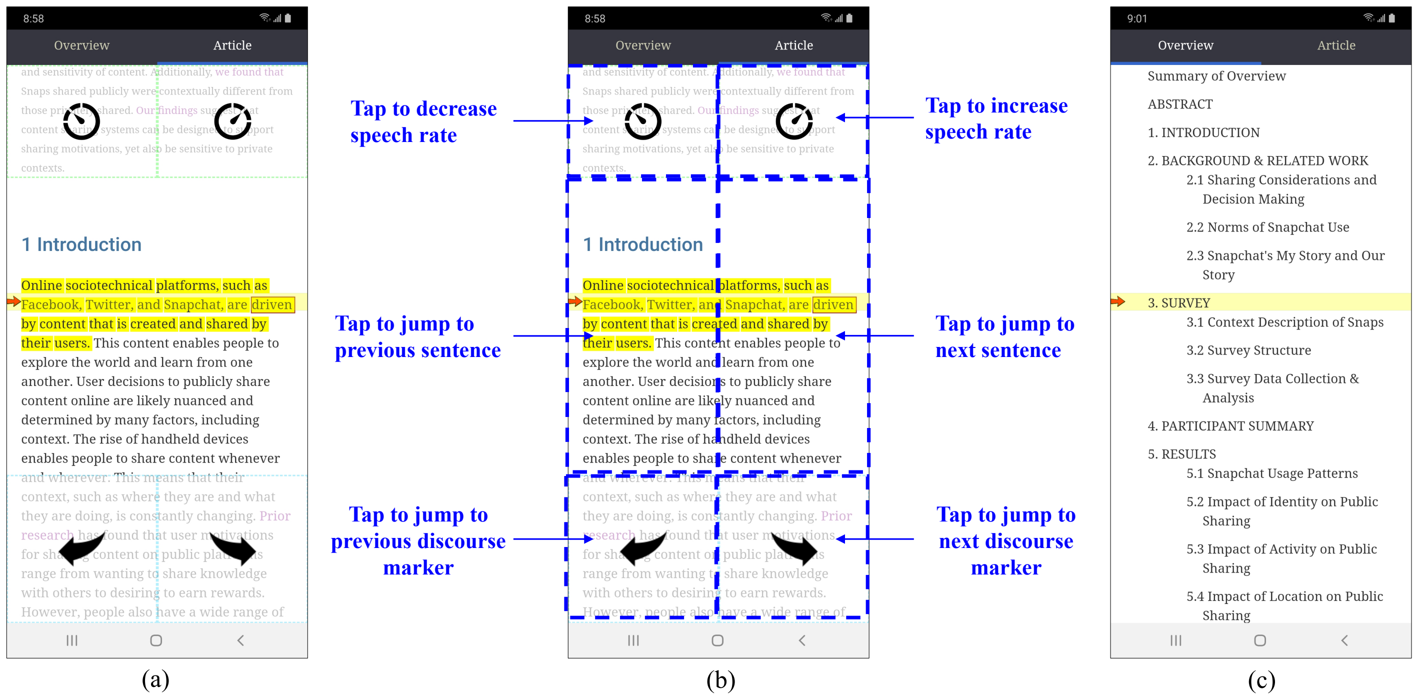


Figure 1.1: (a) Skimmer's user interface (what the user sees). (b) Annotated user interface of Skimmer showing six regions on the screen for different tap-based gestures. (c) Overview page outlines the structural organization of a document. The user switches between the Article page and the Overview page by swiping horizontally.

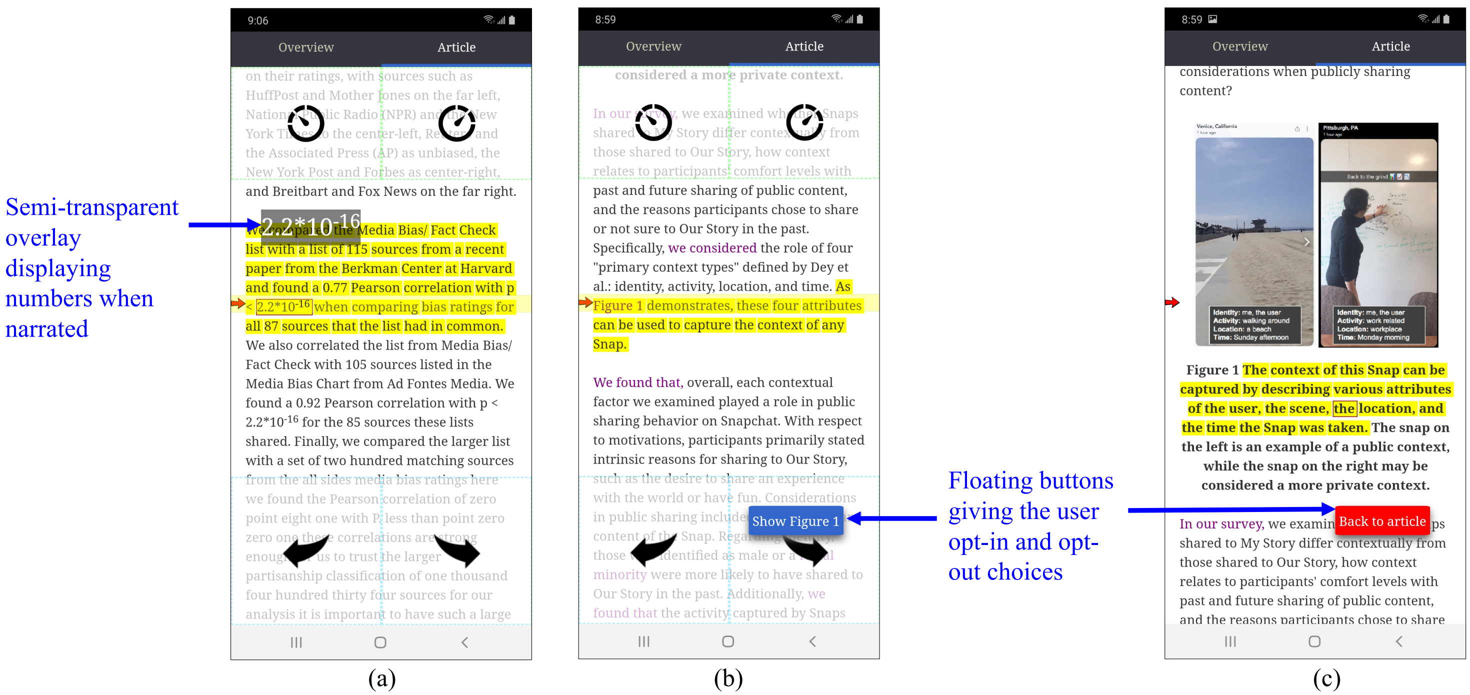


Figure 1.2: (a) Skimmer sets out numbers and reads them in a rounded format to reinforce learning when the user glances at the screen. (b) Skimmer gently informs the user with haptic feedback when a figure/table is referred to in the text. The user has the choice to opt-in (blue button with caption “Show Figure 1”). Upon opt-in, Skimmer shows the image and narrates the caption. (c) The user can resume back to main content at any time by tapping the red button.

To address our research question we:

- conducted a user study (N=20) in a simulated commuting context to understand the needs of auditory skimming,
- iterated extensively on design guidelines for eyes-reduced skimming,
- designed and implemented an eyes-reduced skimming prototype app, Skimmer (Figures 1.1, 1.2), that works with academic documents, and
- conducted a preliminary evaluation (N = 6) on the bus.

1.2 Contributions

Our work contributes the following:

- the concept of multi-modal eyes-reduced skimming that can be used on-the-go,
- design guidelines for eyes-reduced skimming that can be leveraged by other designers, and
- Skimmer, a system that implements those guidelines.

1.3 Overview

Previous work related to this research is summarized in Chapter 2. In Chapter 3, we describe the goal, methods, and results of our need finding study. Chapter 4 outlines our proposed guidelines for eyes-reduced skimming. Chapter 5 illustrates the design of our skimming application. In Chapter 6, we outline the results of a preliminary evaluation of our design. Chapter 7 discusses results and design implications building upon our research. Finally, in Chapters 8 and 9, we discuss the limitations of our research and summarize our research, respectively.

Chapter 2

Related Work

Our study for designing an eyes-reduced document skimming application for situational impairments is informed by an understanding of the notion of situational impairment, eyes-free and non-visual interaction systems, human listening capabilities, audio interfaces, and skimming text and audio content.

2.1 Situational Impairment and Eyes-free Interaction

The proliferation of technology and devices in the lives of human beings have brought about new dimensions in accessibility research. One such example is situational impairment. Situational Impairment (SI) is the temporary effect of factors (behavioral, environmental, attentional, affective, social, technological) on a person's ability to interact with a computing device [13, 14].

Some researchers have studied both situational impairment and mobile device interaction. The study in [15] provides an overview of factors which cause SI and their impact on mobile device interaction. Such factors include - ambient temperature, mobile state, ambient noise, mood and stress, encumbrance, and ambient light. Several studies have explored the impact of an on-the-go setting on mobile device interaction. For example, walking [16], encumbrance [17] have negative impact on mobile device interaction, especially touch accuracy. Several remedies have been proposed in response to these problems, such as improving touch accuracy by generating user-specific touch models [18], enlarging targets [19], CrashAlert [20] for eyes-busy mobile interaction while walking, and eyes-free unimanual bezel-initiated gestures [21].

Understanding needs for non-visual or eyes-free interaction and proposing systems and interaction techniques, especially for reading in SI, have been of interest to researchers. Yi et al. [11] explored users' motivation for eyes-free interaction on mobile devices, and found that certain environmental (difficulty reading from smartphone under extreme lighting condition), social (openly interacting with a phone in a meeting), device specific (small screen causing inconvenience in reading), and personal reasons (self-enthusiasm) motivate people to interact eyes-free. As an example, Ghosh et al. [22] designed EDITalk, an eyes-free interaction for editing word processing (highlight, comment, insert, delete, replace). Furthermore, some studies explored reading in SI. The studies in [23, 24] have shown that walking has a negative impact on reading comprehension. In response, [24] proposed a speech-synthesis based audio interface to improve comprehension while navigating within the environment. SeeReader [25] is an eyes-free document reader for mobile devices. It supports region-based (figures, tables, and paragraphs) navigation by detecting regions under the user's finger while traversing in circular motion, reads out loud summary of regions, while also notifying presence of images. Read4Me [26] is a prototype browser that leverages smartphone sensors to detect context switches and automatically onboards user to a hands-free mode by using TTS. While on the surface, SeeReader and our work may seem similar, in our work, we systematically understand needs and interaction techniques for skimming in SI, and explore how non-linear navigation (sentence, paragraph, section) can support skimming structured documents.

Researchers have also investigated the potential of audio and haptics in eyes-free tasks, such as menu-selection and navigation. Metatla et al. [27] explored audio-only workspace, empirically showing that audio can be used for non-visual collaboration tasks (editing an audio menu). The study in [28] explored eyes-free mobile multi-tasking, showing the effectiveness of spatial audio when used with concurrent audio streams in divided-attention (listening to podcast and browsing audio menu) and selective-attention tasks (listening to music and browsing audio menu). Researchers leveraged audio for eyes-free menu-selection tasks - 3D radial pie menu with head gesture and 2D hand gesture [29], earPod [30], BlindSight [31]. Other studies have leveraged haptic feedback for eyes-free mobile device interaction and menu selection, such as [32, 33]. PocketMenu [34] is a non-visual interaction technique with touchscreen menu items in an on-the-go setting, combining both audio (speech) and haptic (vibro-tactile) feedback. Researchers have also explored eyes-free navigation systems in - (1) digital space - [35] by gesture typing on an invisible and adjustable QWERTY keyboard, and (2) physical space - AudioGPS [36] using spatial audio for GPS navigation. In our work, we leverage both audio and haptics to reinforce action completion in an eyes-reduced context, while doing non-linear navigation for document skimming.

2.2 Human Listening and Audio Interfaces

Researchers have explored the effects of TTS on both sighted and visually impaired (VI) populations. In a recent study, Bragg et al. [37] performed a study on intelligibility of fast, synthetic speech, on both sighted and VI participants. The study showed that VI people can listen to TTS at a higher rate than sighted people. They also showed that sighted people can listen to TTS at a rate higher than the normal speaking rate. Mean listening rate for sighted participants was 297 Words Per Minute (WPM), which is much higher than the typical human speaking rate of 120 - 180 WPM. Cohen et al. [38] recommended guidelines for Voice User Interface design based on the nuances of human listening perceptions. We take important design inspirations from these works, and also put forward our own empirical observations from our study to complement their findings, such as setting an optimum speed for skimming documents.

In order to understand how state-of-the-art read-out-loud interfaces support or do not support auditory skimming, we explored 13 such applications available in the app stores, designed for both visually impaired and sighted users. Most applications support either original layout (which keeps the original text formatting of the document) or a plain text layout (which gets rid off the document's text formatting), or both. Applications differ by types of navigation (spatial, semantic, temporal) and granularity of each types (page, chapter, sentence, 10 seconds, 30 seconds, and so on). We also observed differences in user interface designs - how basic controls are integrated, spatial organization of menus and sub-menus. Figures 2.1, 2.2, and 2.3 summarizes our comparative analysis on how well the existing apps support different features for auditory skimming. The label "Supported" means that the application fully supports a wide range of sub-features that constitute the feature. "Partially supported" indicates an limited feature support. "Not supported" means that the app does not support the features at all.

Our work builds on top of these applications. We use these applications as a first-step to guide our understanding of read-out-loud feature, and shape our understanding of how well eyes-reduced skimming is supported.

2.3 Studies on Skimming

Researchers have explored two categories of source materials for skimming - text document and audio.

2.3.1 Document Skimming

Several studies have explored how people skim a document. Masson [39] defined skimming as a "technique that is commonly associated with reading goals that involve the comprehension of only a subset of a story's content". The reader focuses on information relevant to the goal of skimming. Adler and Van Doren [9] defined skimming as the first sub-level of inspectional reading. According to them, the main aim of skimming is to discover whether the material "requires a more careful reading." Studies such as [40, 3] found that skim reading happens in patches, and the paragraph is the most plausible option as a patch. Therefore, readers move from paragraph to paragraph, sticks to reading a paragraph until rate of information gain drops below a threshold. Marshall and Bly [6] explored within-document navigational patterns in both physical and digital documents. They observed that skimming, scanning, and glancing are common reading strategies among people reading longer documents, embodying *lightweight navigation* such as looking ahead, looking back to re-read for context, and narrowing or broadening focus to an area. Since the widespread use of mobile devices and tablets, researchers have also explored reading behaviors in digital environments, and found evidence of skimming as a popular reading strategy [41, 42, 4, 43]. Our study benefits from the empirical understanding of skimming from the aforementioned studies, confirm their findings through empirical findings of our own, and contributes further understanding of skimming strategies using audio in eyes-reduced manner.

Researchers have proposed novel interfaces to aid document skimming. Structure-Aware Touch-Based Scrolling (SATS) [44] assists the reader to perform non-linear navigation. In SATS, the tablet screen is divided into four parts vertically, assigning each part to navigational units, such as chapter, section, sub-section, and page. Spotlight [45] is an attention-optimised skim reading tool which selects salient objects (e.g., title, headings, a figure) and displays as a transparent overlay on the screen as the user scrolls. In our work, we take advantage of gesture-based navigation to support eyes-reduced skimming. In addition to Spotlight's findings, we found that discourse markers are also salient objects, grounded from our first study in Section 3.

2.3.2 Audio Skimming

SpeechSkimmer [5] is a speech-based skimming device. In SpeechSkimmer, users can skim recorded speech at different levels by - time compression, pause shortening and structure identification (pauses inferring sentence boundaries, pitch change inferring speaker change). SpeechSkimmer also uses non-speech sounds to provide navigational cues. SCAN [46] and [47] support audio skimming by using transcripts as proxy to visually skim audio content and keyword spotting by confidence shading respectively. Recent studies focused on non-visual skimming for screen reader users. Studies in [48, 49] support skimming online contents by automatically summarizing texts in multile levels. They extended this study to touch screen devices [50], and built on top of VoiceOver's default gestures, such as pinch-in and out to control levels of summarization. In our work, we extend the literature by contributing Skimmer. Skimmer uses non-linear navigation exploiting semantic structure of long, structured documents. More importantly, it is designed to be used largely eyes-free by enabling the user to predominantly listen to the text, but also supports selectively glancing at the screen, should the user choose to do so, an interaction technique we termed as "eyes-reduced" interaction.

Application	Structural organization	Localization and spatial awareness	Navigation: Units of navigation	Navigation: Interaction technique	Navigation: Keyword or phrase spotting	Skipping over footnotes, citations, page headers	Differenciating meta and main information	Referring Figures/ Tables	Speech rate control	Annotation (highlight and bookmark)	Document layout
Skimmer	*Supported.* An Overview page is available. Does not require visual attention since items are narrated.	*Supported.* Highlights current word and sentence. On-demand navigation to Overview page for spatial awareness.	*Supported.* Sentence, paragraph, sub-section, section, discourse markers.	*Supported.* Zone-based taps and standard touch gestures.	*Supported.* Discourse markers support navigating to keywords or phrases.	*Supported.*	*Supported.* Low-pitch and high-pitch voice for meta and main information respectively.	*Supported.* Informs user by a haptic vibration, shows figure/table if user opts-in.	*Supported.* Eyes-free tap gesture.	*Not supported.*	Original layout in HTML format.
VoiceDream Reader	*Partially supported.* A table of content is available, requires visual attention because items are not narrated.	*Partially supported.* Highlights current word and sentence. No feature for spatial awareness.	*Supported.* 15s, 30s, 60s, sentence, paragraph, page, bookmark. Sentence and paragraph do not work reliably.	*Supported.* Soft-buttons at the bottom of the screen. Mode switching required to activate different navigation units.	*Not supported.*	*Supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* 1. Two finger swipe up/down to increase/decrease. Does not work in original layout. 2. Slider under a menu at the top of the screen.	*Partially supported.* Highlight a sentence. Optionally add notes. Notes are not narrated by voice.	Supports both original layout and plain text layout.
@Voice Aloud Reader	*Not supported.*	*Partially supported.* Highlights current line.	*Partially supported.* Sentence.	*Partially supported.* Soft-buttons at the bottom of the screen.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* A sliding tray at the bottom reveals slider control for speech rate.	*Not supported.*	Plain text layout.
Amazon Kindle and Fire Tablet	*Partially supported.* A table of content is available, requires visual attention because items are not narrated.	*Partially supported.* Shows percentage of completion.	*Partially supported.* 30s, page.	*Partially supported.* Soft-buttons for 30s, tap on either side of the screen for page navigation.	*Partially supported.* X-ray feature shows all important entities in the article.	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Tapping a soft-button reveals speech rates: 1x, 1.5x, 2x, 3x, 4x.	*Partially supported.* Highlighting and thumb bookmarking to switch between two pages.	Original layout of the document.

Figure 2.1: (Part 1) A comparison of the level of support for various skimming features in different read-out-loud apps. Our design, Skimmer at the top row, supports all features except annotation. VoiceDreamReader (partially) supports most features and hence was used in our need finding study (Chapter 3) and preliminary evaluation (Chapter 6). The rest of the applications are presented in an alphabetical order. Green, yellow, or red background depicts whether features relevant to supporting skimming are 'supported,' 'partially supported,' or 'not supported'.

Application	Structural organization	Localization and spatial awareness	Navigation			Skipping over footnotes, citations, page headers	Differenciating meta and main information	Referring Figures/ Tables	Speech rate control	Annotation (highlight and bookmark)	Document layout
			Units of navigation	Interaction technique	Keyword or phrase spotting						
Capti	*Not supported.*	*Partially supported.* Highlights current word.	*Partially supported.* Word, sentence.	*Partially supported.* Soft-buttons on the screen.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Tapping soft-button reveals speech rate control slider.	*Not supported.*	Original layout of the document.
iBooks with VoiceOver	*Partially supported.* Page thumbnails available.	*Partially supported.* Highlights selected line momentarily, but does not persist.	*Partially supported.* Letters, words, and lines.	*Partially supported.* Depending on mode of navigation, flick up/down to move back/forward.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Similar to basic navigation, switch to speech rate mode.	*Not supported.*	Original layout of the document.
NaturalReader	*Not supported.* 'Chapter' feature available, but does not work with PDFs.	*Partially supported.* Highlights current sentence.	*Partially supported.* Units of navigation is unclear. Probably temporal navigation is supported.	*Partially supported.* Soft-buttons at the bottom of the screen.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Slider control hidden in two sub-menu levels.	*Not supported.*	Supports both original layout and plain text layout.
PDF Voice Reader	*Not supported.*	*Not supported.*	*Partially supported.* Page.	*Partially supported.* Scroll to next or previous page.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Tapping a soft-button reveals speech rates: slow, normal, fast.	*Not supported.*	Original layout of the document.
Pocket	*Not supported.*	*Not supported.*	*Partially supported.* 15 seconds.	*Partially supported.* Soft-buttons at the middle or bottom of the screen.	*Not supported.*	*Not supported.*	*Partially supported.* High pitch voice reads out article title only, low pitch voice reads verbatim content.	*Not supported.* Continues reading the main text.	*Supported.* Two taps required to show increase/decrease option.	*Partially supported.* Highlight by selection.	Plain text layout.

Figure 2.2: (Part 2) A comparison of the level of support for various skimming features in different read-out-loud apps.

Application	Structural organization	Localization and spatial awareness	Navigation			Skipping over footnotes, citations, page headers	Differenciating meta and main information	Referring Figures/ Tables	Speech rate control	Annotation (highlight and bookmark)	Document layout
			Units of navigation	Interaction technique	Keyword or phrase spotting						
ReadAloud: Voice Reader	*Not supported.*	*Partially supported.* Highlights current word.	*Partially supported.* Sentence.	*Partially supported.* Soft-buttons at the bottom of the screen.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Tapping a soft-button reveals speech rate control slider.	*Not supported.*	Original layout of the document.
Read Out Loud Adobe Reader	*Partially supported.* Page thumbnails and section title outline is availalbe, requires visual attention because items are not narrated.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Not supported.*	*Not supported.*	Original layout of the document.
Select and Speak	*Not supported.*	*Partially supported.* Momentarily highlights current line after user selection.	*Partially supported.* Sentence.	*Partially supported.* Soft-buttons for sentence navigation.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Partially supported.* Available in pro version only. Speech rate control available under a menu item.	*Not supported.*	Original layout of the document.
TalkBack Android	*Not supported.*	*Partially supported.* Highlights current word.	*Partially supported.* Word, sentence, section title.	*Partially supported.* Soft-buttons at the bottom of the screen.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Soft-buttons at the bottom of the screen.	*Partially supported.* Depends on the PDF reader used.	Depends on PDF reader used.
Voicepaper	*Not supported.*	*Partially supported.* Highlights current word.	*Partially supported.* Page.	*Partially supported.* Scroll to next or previous page. Tap to start reading from a particular place.	*Not supported.*	*Not supported.*	*Not supported.*	*Not supported.* Continues reading the main text.	*Supported.* Two taps required to show increase/decrea se option.	*Partially supported.* Bookmarking supported.	Original layout of the document.

Figure 2.3: (Part 3) A comparison of the level of support for various skimming features in different read-out-loud apps.

Chapter 3

Understanding User Needs for Auditory Skimming

We conducted a formative study to understand the needs for auditory skimming in situationally impaired context. In particular, we sought to understand how might people skim auditorily, what are the challenges, what is working and not working, given the state-of-the-art read-out-loud applications (hence referred to as "Need Finding Study" elsewhere in this thesis). We conducted the study in a controlled lab setting designed to mimic some of the conditions of a public transit bus. We used a representative read-out-loud app as a probe. All user study materials for this study can be found in Appendix A.

3.1 Method

3.1.1 Participants

We recruited 20 university graduate students (10 male, 10 female) through convenience sampling. Participants came from various backgrounds (e.g., Computer Science, Kinesiology, Occupational Therapy, Business). We recruited only those who reported being familiar with reading research articles, because unfamiliarity or inexperience might introduce bias in the results.

3.1.2 Task

Participants were asked to imagine that they are enrolled in a graduate-level course and have reading assignments for a class later that day in which they have to participate in a peer discussion session. In order to contribute to the discussion, they should skim auditorily under a time constraint while riding a bus to school. We asked participants to minimize visual attention to the screen, pretending to experience motion sickness. We also mentioned that it was acceptable to look at the screen while issuing a command (e.g., increasing speech rate, or tapping a button). However, we stressed that they should try to look away when listening. We told participants that their objective was to gain an understanding of - gist of the article, topic and problem statement, key takeaway, and structural organisation of the article. Our intention for participants was not that they give us ”the right answer.” Rather, we gave them an objective in order to induce them to try their best. Specifically, we told them that they would be asked the above-mentioned objectives during post-task interviews.

3.1.3 Design of the Study

Our study was a 2 (document formats) vs. 2 (article types) within-subjects design. In total, participants completed the task 4 times, once in each condition. The order in which participants saw document formats and article types was fully counterbalanced.

3.1.4 Conditions

Document formatting condition: Based on our empirical findings from existing read-out-loud applications, we found that most applications support either one or both of the following layouts - plain text layout (displays only text and discards

Article	# Words	Reading ease level	Time allocated
A1 [51]	8202	College graduate	8.5 mins
A2 [52]	7260	College graduate	8 mins
A3 [53]	4664	College	4.5 mins
A4 [54]	4559	College	4.5 mins

Table 3.1: Articles (in rows) selected for skimming task and properties (in columns) that defined inclusion and exclusion criteria for skimming task

any formatting) and original layout (displays the original formatting condition of the document). We studied both of them to understand if one or both are favorable to auditory skimming.

Article type condition: We chose documents that are isomorphic in terms of topic, structure, and difficulty. For this study, we chose the following two article types - research paper (representative academic papers A1 and A2 in Table 3.1) and Pew report (representative professional reports A3 and A4 in Table 3.1). A summary of the selected articles is given in Table 3.1. For reading ease level, we measured Flesch-Kincaid Reading Ease [55]. We followed the reading material selection rationale mentioned in [40]. 'Social media' was the common theme among all four articles. In addition, having 'college' or 'college graduate' reading difficulty level ensured that the articles were suitable for university graduate students, regardless of their academic backgrounds.

We imported PDF files of the articles in the reading app. All articles were single column typeset to maintain consistency. Before that, we fixed reading order using Adobe Acrobat Reader's accessibility tool.

3.1.5 Apparatus

We used an iPhone 5S device for this study. We explored 13 read-out-loud applications in both mobile and desktop platforms (in Figure 2.1, 2.2, 2.3). We chose to use VoiceDreamReader (VDR) as a technology-probe for the study because of its comprehensive feature sets and positive ratings by its users in app stores. We demonstrated the following navigational features of VDR to the participants

Figure 3.1: Participant standing in front of a big screen where we projected ambient bus scenario.

- play/pause, scrolling, jump forward or backward by 15 seconds, double tapping to start reading from a place, and page jumps.

VDR provides feature support for non-linear navigation such as sentence, paragraph, sections, however, these features did not work consistently for all 4 documents. The temporal navigation worked reliably, therefore, we decided to use temporal navigation (15 seconds) only.

3.1.6 Procedure

Participants at first answered demographic questions (age, gender, discipline). Then we introduced VDR to the participants. We demonstrated the features and asked them to try them out. To put participants in a bus riding context in a controlled lab setting, we projected a video that captured window view from a bus commuter's perspective on a large screen, with high volume to emulate ambient noise (in Figure 3.1).

We determined the task time given for skimming an experimental material (see Table 3.1) by adopting Duggan et al.'s approach [40] that calculates the ratio of number-of-words in the material and the target reading speed for skimming. In our study, we used the human *listening* rate as a base to determine the task times. Bragg et al. [37] reported that average listening rate for sighted people is 297 words per minute (WPM). Our pilot tests showed that it was difficult for people to comprehend when listening at 297 WPM. Setting the reading time following this protocol would result in very long reading time (27 minutes for A1), which could hint to participants to listen linearly and not skim it. To solve these problems, we conducted more pilot tests at varying listening speeds, such as 180 WPM, 200 WPM, 225 WPM and 250 WPM. We decided to set 225 WPM as default as pilots agreed that it is a reasonable speed to support their listening comprehension. Also, participants were free to change the reading speed as needed. Based on pilot tests, we found that setting 8.5 minutes for A1 and 4.5 minutes for A3 was reasonable, and thus we set the time for the rest of the articles to be commensurate with A1 and A3.

After each reading, they answered perceived level of comprehension. Participants verbally responded to our questions related to skimming objectives. During the task, participants saw either original layout or plain text layout first. After each layout, they answered NASA TLX [56] workload measurement for the task.

3.1.7 Quantitative Measures

We measured the following: (1) NASA TLX - a six scale weighted measurement for workload, and (2) perceived level of comprehension - a 5-point Likert scale measuring their agreement on how well they understood each skimming objectives (see Appendix A.4).

3.1.8 Data Collection and Data Analysis

In addition to quantitative measures, we collected qualitative measures such as observation data, video record of the tasks, and audio record of post-task interviews. In the post-task interviews, we first asked questions about the gist and structural organization of the article to gauge their understanding of the article. Apart from conventional questions pertaining to what worked well and did not work well, we also asked for more details on what, why, and how of particular instances. For analysing qualitative data, we maintained a codebook. We started line-by-line coding, trying to explain the data, looking at the actions, and their outcomes. The initial round of coding was in-vivo [57]. We kept memos defining the codes. Codes and memos were shared and discussed with the research team, and refined altogether. We analyzed quantitative data using repeated measures ANOVA and Aligned Rank Transform [58] because of non-parametric multi-factor data in our analysis.

3.2 Findings

The central theme from our qualitative observation is that **the pull to look at the device is very strong.** Participants almost seemed compelled at times to look down when skimming. Analysis of the video footage shows that participants were visually "scanning", which they confirmed in the post-task interview. They indicated that they were visually foraging parts of text where they believed the gist of the paper's content might be (e.g., end of Introduction, and places marked by phrases such "In this paper" or "Our contributions are" might appear.).

This tendency to see the document was so strong that four participants would not follow the study protocol requesting them to minimize their visual attention to the screen. They read-along most of the time, meaning that they visually read the text while just following the auditory narration as a secondary information channel. Individual preferences seem to play into such behavior as some expressed themselves as a "visual person" and explained that "[listening to] audio is not learning." (P10) We discarded their quantitative data to keep homogeneity in our analysis (below) but kept their qualitative data, as their responses add accounts for why the pull to look is so strong.

Overall, the participants struggled between using two different modes of reading: eyes-free vs. visual reading. Skimming, as a particular type of reading behavior, often requires frequent non-linear navigation and occasional glancing over to visual content, such as figure. However, the traditional read-out-loud apps, including VDR, are designed with the assumption of a linear consumption of audio. This challenge calls for us to design and develop a new auditory skimming app motivated to support a new mode of text consumption: *eyes-reduced* reading that flexibly blends auditory and visual modes of reading. The following findings specify what are the concrete user needs and challenges for supporting skim-reading in a eyes-reduced way.

Non-linear navigation is hard. Participants were in need of non-linear navigational features that respect the semantic structure of the document (e.g., sentences, paragraphs) so they do not feel lost right after the jump, which is a frequent navigational pattern in skim reading. They did not find *temporal* jumps (10s, 15s, and so on) useful, because *"it wasn't clear where 15 seconds leap would take me to. Move by sentences is more meaningful."* (P8) In general, participants reported being familiar with the structure of research papers and what to expect from them. *"I wanted to navigate. I wanted to see the results, see the summary, to the suggestions. But it was not easy to navigate to them. You have to scroll and find."* (P4) Participants wanted to listen to the first one or two sentences of a paragraph and skip to next paragraph. However, they often did not skip because *"double tapping the next paragraph was not working always."* (P8) In addition, it required visual input from them to make a selection. Navigation was more difficult with plain text layout because *"there was no structure and everything looked similar"* (P16) and participants had to *"scroll and scroll until at some point I became frustrated."* (P9)

Listening and navigating at the same time is difficult. The traditional app keeps narrating the text even after when the user starts the navigational interaction (e.g., scroll to find where to jump into). Participants found this simultaneous stimulus cognitively taxing. It is known that listening and short-term memory is competing for the shared cognitive resources [59]. Their work-around was to pause the audio before navigating although cumbersome.

Some types of content are difficult to understand by listening. Some participants preferred *looking down* at a figure or a table referenced in the text, because they were visual by nature: *"It's useless to say Figure 2 when you can't actually see Figure 2."* (P7) Some of the *textual* content was also difficult to understand, even when the narration is reading it verbatim. Such items include complex numbers, and acronym: *"[It was] difficult to understand when the voice is telling p value is 0.3."* (P7) *"[It was] wordy, the stat reports are about numbers, and hearing the voice reading the numbers was not useful."* (P13)

Original layout better supports structural understanding. For understanding the layout or structure of the article, participants rated that they understood better with original layout formatting ($M = 3.81, SD = 0.44$) than plain text layout formatting ($M = 2.38, SD = 0.35$); $F(1,15) = 28.31$, $p < .001$, partial η^2=.65. In several instances, participants reported the benefits of original layout over plain text layout, such as *"the original layout was better. If the purpose is to skim the article, I would like to see the layout"* (P4) and *"I like the original layout more, because it was impossible to tell how the paper is laid out in the plain text view."* (P7) While doing the task, workload was significantly higher in plain text layout ($M = 76.67, SD = 11.87$) than using original layout ($M = 63.97, SD = 15.22$); $t(15) = -4.42$, $p < .001$.

Dynamic speech rate control is needed but inaccessible. Participants reported that they preferred slow narration for important content and faster narration for trivial content. Such selective reading pattern is similar to what is reported in the literature on visual skim reading behaviors [3]. However, participants often could not increase or decrease the speech rate, even when they wanted to, because the rate control was not accessible, buried deep in the menu. *"But since every time*

I have to touch it and decrease and increase the word speed, this deviated me from what I was hearing. That's why I skipped increasing the speed later." (P12) Moreover, setting a default speech rate was another barrier, because the unit of speech rate "words per minute" was not meaningful for estimating how fast (or slow) it would be. Therefore, participants listened to the audio to figure out which speed to set.

Individual differences observed in reading order. We found reading orders for research papers varied among participants. This order seemed predetermined for a person, something they learned and refined over their experience of reading research papers. For example, the order of Abstract, Introduction, and Conclusion was very popular. Some would go to the Discussion section depending on how well they utilised their time. Only P16 and P19 said they would be interested to see the figures first. For Pew reports, which were relatively new to many, they either (1) listened chronologically, (2) visually read section, sub-section titles and listened to one or two paragraphs, or (3) listened to the first paragraph and last paragraph of each section.

Document formatting and article type impact comprehension. When it comes to understanding the topic, people perceived to have better understanding - (1) while using original layout ($M = 4.16, SD = 0.13$) than plain text layout ($M = 3.53, SD = 0.31$); $F(1,15) = 7.147$, $p < .05$, partial η^2=.32, and (2) when reading research papers ($M = 4, SD = 0.35$) than pew reports ($M = 3.69, SD = 0.53$); $F(1,15) = 5.43$, $p < .05$, partial η^2=.27. For understanding the key takeaway, participants perceived to have better understanding (1) while using original layout ($M = 3.31, SD = 0.80$) than plain text layout ($M = 2.63, SD = 0.53$); $F(1,15) = 11.84$, $p < .05$, partial η^2=.44, and (2) when reading research papers ($M = 3.44, SD = 0.62$) than pew reports ($M = 2.5, SD = 0.35$); $F(1,15) = 14.27$, $p < .05$, partial η^2=.49.

Create annotation for later consumption. Part of the motivation for skimming is to decide whether the document needs detailed reading later. To this point, participants mentioned their interest to create annotations such as, highlighting, putting starts, and bookmarking. Another motivation is to mark which content is important and go back to it sometime later.

Skimming in other situationally impaired contexts. Participants imagined other situationally impaired contexts where they could find auditory skimming useful. For example, P17 and P19 said they would skim while doing regular daily activities, such as cooking, doing exercises. P11, a Kinesiology student said she might skim an article auditorily during outdoor activities, as pulling up a laptop is not always convenient. P16 and P20 mentioned that auditory skimming can be helpful while waiting in a queue (e.g., boarding a plane).

To summarize our findings from this study, auditory skimming required 'eyes-reduced' reading rather than entirely eyes-free. Participants blended visual (for content that is predominantly visual, such as images) and non-visual modes of reading in order to skim using audio. Specifically, non-linear navigation was key to successful skimming, but it is not supported well.

Chapter 4

Design Guidelines for Eyes-Reduced Skimming

We triangulated implications from multiple data sources to develop a set of robust and comprehensive design guidelines. The data sources include: (1) the results from the need finding study in Section 3, (2) existing theoretical and empirical studies of skim reading [3, 6, 9, 39], and (3) existing guidelines for creating voice narrations and voice user interface design guideline [38, 60]. Here we present our design guidelines:

DG-1. Provide ways to navigate the structure of the article in a non-linear fashion and to localize the current position. This should be available to the user on demand, similar to using a Table of Contents. The primary support required here is the ability to jump to sections and to easily know the current position in the document.

DG-2. Provide spatial and semantic navigation instead of temporal navigation. Temporal navigation did not help the user to conceptualize their jump location. Scrolling as a spatial navigation is very common. Hence, it should be supported. Semantic navigations are associated with the meaning of words and underlying structure of the article. Hence, there should be support for sentence and paragraph level jumps. Moreover, informed by the need finding study and discourse analysis on academic articles by Hyland and Tse [61], we recommend supporting discourse marker jumps to facilitate skimming since they also act as important cues when skimming visually. The study in [61] illustrated two major types of discourse markers - interactive and interactional. Interactive are those that writer uses to manage information flow as a guide to the reader, such as 'finally', 'to conclude', 'in this paper', 'see Figure 1'. Interactional markers are those that convey the author's perspective towards the propositional information. For example, 'it is clear that', 'note that', 'our contributions'.

DG-3. Pause narration when the user is navigating. Narration should be paused when the user is navigating, such as scrolling or dragging. This is because while navigating, the user focuses on where to go next and does not necessarily focus on current narration.

DG-4. Provide ways to adjust speech rate dynamically. The user should be able to adjust the speech rate dynamically without pausing or stopping audio. This option should be accessible to the user without having to go at sub-menu level hierarchy. Moreover, a simple scale should represent speech rate, e.g., 1.0 for normal, 2.0 for twice the normal speed, instead of 'words per minute'.

DG-5. Provide ways to refer back visually to text content from the auditory narration and vice versa. A spatial point of reference can help the user to refer back to text content at any moment while skimming. There are two ways to support this. First, the spatial point of reference can remain static, and the content can 'flow' around the static point. An example is to have a static point of reference mid-way down the smartphone screen along the left bezel, leaving the text content scroll-able. Second, the spatial point of reference can move across static content. The user should be able to move the point of reference that will serve as the play head of the narration.

DG-6. Provide rich paralingual cues in narration for better listening experience. For a better listening experience and to motivate understanding, the speech synthesizer should support rich paralingual cues in narration. Some examples are - breakdown lists into groups of 3 or 4, narrate statistical values following verbal reporting conventions, expand abbreviations (‘e.g.’, ‘i.e.’), insert breaks in long sentences to help user absorb information, avoid footnotes and citations in the text. Moreover, the system should add *context* to help the user focus. For example, announcing which section is going to be narrated next.

DG-7. Provide auditory or haptic feedback as non-visual confirmation. For non-visual affordances, auditory or haptic feedback should be given. For example, an auditory feedback can be given between end of a paragraph and start of a new paragraph, to nudge the user to pay attention to an important content, such as an image or a number.

DG-8. Support opt-in visual engagement. Support opt-in or consensual processes when switching context (e.g., jumping from main text to figure/table caption). Also, set out numerical texts from other texts when narrated to reinforce learning when the user glances at the screen.

DG-9. Support unimanual interaction for eyes-reduced skimming. Generally speaking, user interface widges, such as menus and buttons, should be easily reachable, assuming that the user will be primarily interacting with one hand.

DG-10. Support individual differences in skimming strategies. There should be support to preset a reading order and automatically listen through the text without the user explicitly navigating to parts of the text. This can happen when the document type/format is familiar to the reader and knows what to expect. Furthermore, there should be support for choosing which part to listen to on the fly. This can happen when the user is not familiar with the document.

DG-11. Support annotation creation and consumption. The user should be able to create within-text telegraphic annotation [62], such as highlight, bookmarks, or putting stars. User should be able to navigate to these annotations. An advanced support would be to create speech annotations. At the same time, there should be support to consume these annotations.

In summary, we proposed a set of design guidelines for eyes-reduced skimming for situational impairments. We formulated these guidelines following our need finding study and our empirical and theoretical understandings of skimming, voice narration, and voice user interface design. Our hope is that other designers will also benefit from, and iterate and refine these guidelines.

Chapter 5

Design of Skimmer

We built a novel read-out-loud app called Skimmer that embodies the majority of the design guidelines (Chapter 4) for eyes-reduced skimming. Skimmer features navigational support for auditory and visual modes of reading, non-linear navigation, and auditory and haptic cues. In this chapter, we explain the design of Skimmer first, and then we explain the iterative design process.

5.1 Eyes-Reduced Skimming Features of Skimmer

Here, we elaborate on how Skimmer supports eyes-reduced skimming for situational impairment and our rationale behind its design.

5.1.1 Basic Navigation for Eyes-Reduced Skimming

Skimmer predominantly supports eyes-free navigation. The user can skip sentences or re-listen to a missed passage with a lightweight, eyes-free tap gesture on the right or left side of screen respectively (Figure 1.1 (b)). Skimmer also supports sentence-level navigation using vision by traditional scrolling (DG-2). When skimming, this can be useful for localizing one's reading position–peeking at how far one is in the text and checking the remaining length of passages. Skimmer also pauses the audio as soon as the user touches the screen for navigational action (DG-3). Paragraph navigation is done through up-or-down vertical flicking gestures. This is because scrolling is so intuitive to people and the content is flowing vertically, so Skimmer aims to promote positive transfer from visual scrolling to eyes-free gestural interactions.

5.1.2 Designing for Easy to Understand Audio Narration

We designed the way Skimmer narrates the written content to help listeners better understand complex and difficult texts: (1) Written text and structural information are narrated by two different voices to help distinguish meta-information from verbatim content. This is done by using a low-pitch voice for narrating written text and a higher-pitch voice for narrating meta-information. The meta-information is for providing context to the reader, such as “Section 1. Introduction”, where “Section” is the meta-information. (2) Skimmer breaks down long lists of items or bullets in groups of 3 or 4 items at a time, to help memory retention (DG-6). (3) Skimmer reads numbers in a rounded format (such as $p < 2.2 \times 10^{-16}$ narrated as $p < .001$), because complex numbers are hard to understand just by listening (DG-6). Skimmer also provides an overlay to reinforce learning as the user visually engages with the numbers (Figure 1.2 (a)). (4) To catch the user’s attention to a discourse marker, Skimmer narrates it louder and slower. For example, for the given text “In this paper, we study Snapchat ...,” Skimmer can identify “In this paper,” as a discourse marker, and add an <emphasis> tag of Speech Synthesis Markup Language (SSML) around it so our TTS engine can emphasize the phrase with a salient pitch and tone. For this, audio narrations in Skimmer were generated using a Google Wavenet-based TTS engine that supports SSML tags.

5.1.3 Navigating Between Discourse Markers

To help users find and navigate to important nuggets of information, Skimmer supports selective jumps between various discourse markers (DG-2) through eyes-

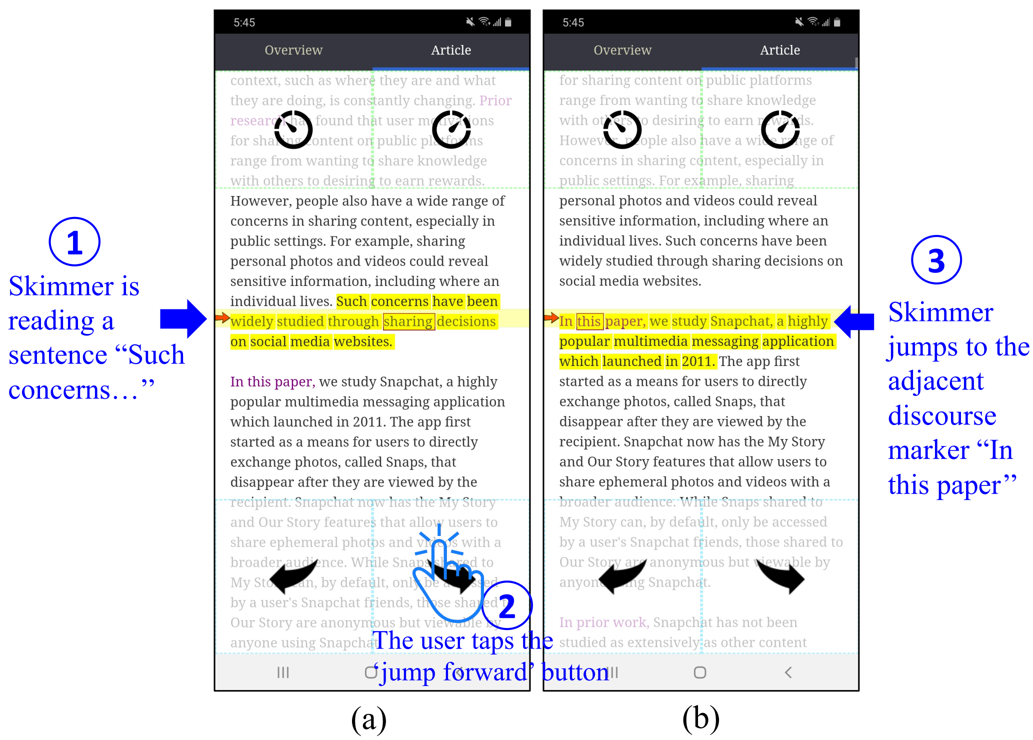


Figure 5.1: Skimmer’s discourse marker navigation. (a) Current reading position of Skimmer. The user taps on the bottom right button. (b) Skimmer jumps to the next discourse marker. In this case, it is “In this paper”.

free tap gesture (lower end of screen, Figure 5.1 (a)). When the user taps the discourse marker navigation button on the bottom right (left), Skimmer moves the play-head to the next (previous) adjacent discourse marker, narrates the discourse marker first (e.g., "In this paper"), and then narrates the full sentence from the beginning of the sentence for context (Figure 5.1 (b)). Both discourse markers and full sentences are content that is narrated by the same voice (low-pitch voice). To reinforce to the users that they have reached a discourse marker narration, Skimmer plays a subtle ambient earcon (a cricket sound) as background audio.

5.1.4 Understanding the High-level Organization

To support structural glancing and navigation through high-level structure of the document (DG-1), Skimmer features a table-of-content style *Overview* page different from the main document view (Figure 1.1 (c)). All content in Overview is accessible eyes-free, as the allowed gestures between the main document view and Overview view are consistent.

5.1.5 Opt-in to See Visual Content

Skimmer's design for eyes-reduced reading nudges users when they encounter visual content. When a narration content includes a figure or table referred to in the text, Skimmer gently informs the user with haptic feedback so that they can look at the visual content (DG-7, DG-8). Skimmer facilitates navigating back and forth between the main text and the visual content on-demand; note that we designed the transition to visual peeking to be *opt-in*; allowing the user to choose to continue focusing on listening to the text or to jump to the visual content (see Figure 1.2 (b)). The haptic vibration pattern was carefully chosen. From Seifi et al.'s [63] library of vibration patterns, we chose the five patterns that have at least one of two relevant tags: "attention-catching" that catches the user's attention when there is a figure/table reference and "get ready" that urges the user to be ready to see the figure/table. Among the five, we picked the one with the highest pleasantness score.

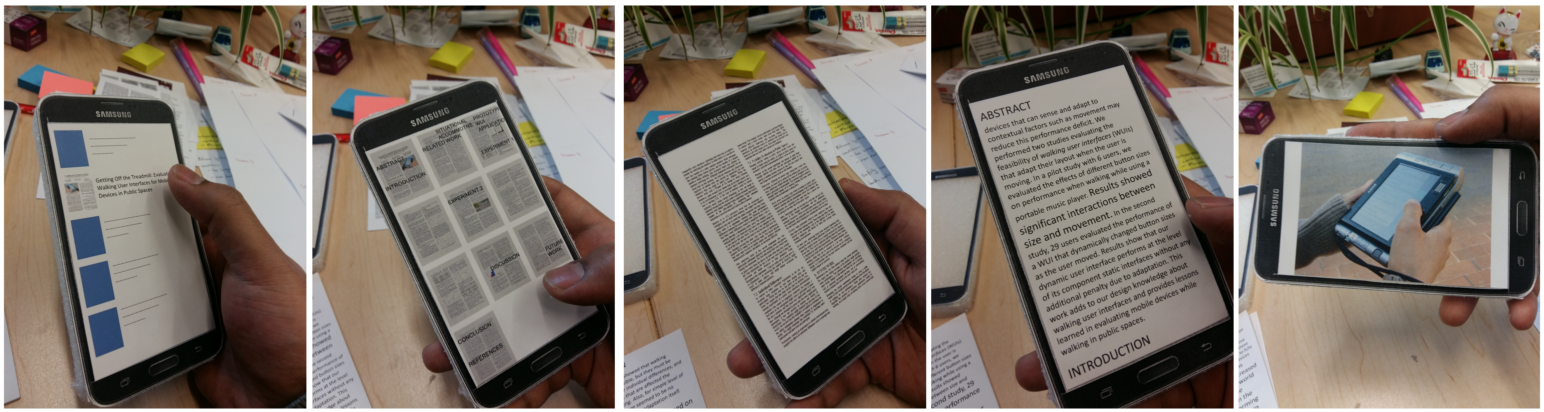
5.1.6 Auditory Cues in Support of Skimming

Drawing inspiration from SeeReader [25], throughout all these features, Skimmer leverages a limited set of distinctive earcons to provide subtle auditory cue/feedback (DG-7). It's worth noting that we consciously limited the number of total distinctive earcons to be under 6, as we found that incorporating too many earcons can cause overlaps of aural characteristic between them, which in turn confuses users. Hence, we used earcons, which is a scarce design resource due to the limited number we can leverage. The purpose of using earcons is to signify the system status changes at the critical path of skimming activities, such as switching between sentences, paragraphs, and discourse markers, and navigation to Overview page–thus giving users clear feedback that Skimmer has processed their navigational gestures.

5.1.7 Accommodating Individual Differences in Listening Rates

Skimmer accommodates individual differences by enabling the user to dynamically adjust the speech rate through an eyes-free tap gesture (DG-4). Skimmer provides two buttons at the top, similar to the discourse marker buttons at the bottom in style and size (Figure 1.1 (b)). Although the top of the screen is often hard to reach, we made a design trade-off by prioritising discourse marker navigation, an oft-used navigational feature, over the rate changes.

In summary, Skimmer embodies our proposed design guidelines to support eyes-reduced skimming. Skimmer implemented the majority of the design guidelines (DG-1 to DG-9). In Chapter 7, we discuss the rest of the design guidelines and our future work.



(a) (b) (c) (d) (e)

Figure 5.2: (a) Initial screen that lets the user choose a document for skimming. (b) The next screen shows thumbnails of every page with section names appearing as overlays. (c) Default layout showing the double column format - the original layout of the document. (d) When the user engages with the text by looking at it, it reflows the text by zooming in on the current text segment. (e) When a figure/table is referenced, the application shows the image in landscape orientation to provide a better view.

5.2 Design Process

We designed and developed Skimmer through a three-phase iterative design process. The first phase was to outline conceptual designs. We ideated several distinct conceptual models that abide by the guidelines, and then created and evaluated low-fidelity prototypes. In the second phase, we refined our low-fidelity prototype to simulate eyes-reduced skimming using a wizard-of-oz technique. Feedback from this simulation motivated us to step into the next phase, where we built a usable and working mobile app to evaluate our design ideas and design guidelines.

5.2.1 Conceptual Design

The first step of our design process was to create conceptual designs of a basic spatial layout and gesture interactions.

Designing the spatial layout

Design ideas for the spatial representation are grounded to findings from Chapter 3 where we learned that original layout is preferable over plain text layout. Therefore, to flesh out our conceptual ideas into tangible design objects, we created a low-fidelity prototype using styrofoam and paper. In our first design iteration, we considered the original layout of the source document (e.g., double column format), as shown in Figure 5.2 (c). We borrowed the concept of space filling thumbnails [64] to provide a high-level structural idea of the reading material to the user (in Figure 5.2 (b)). When the user chooses a section to read, the app displays the page containing that section (Figure 5.2 c). Since our design motivation was eyes-reduced interaction, having a double column layout on the screen with small fonts was a design trade-off that we thought would also discourage the user from focusing on the screen for a long time. However, when the app detects that the user is focusing on the screen, the app reflows the text by joining both columns chronologically and zoom-in to the text segment currently narrated on the screen, as shown in Figure 5.2 (d). Moreover, since documents such as research papers come with content that require visual attention, such as figures or tables, our design lets the user to visually engage with the figure or table once they are referred in the main text by displaying in landscape orientation on the screen (in Figure 5.2 (e)).

Mapping features to gesture sets

Our target at the outset was to optimize the touch gesture vocabulary in such a way that users can positively transfer easy and well-known interaction techniques to Skimmer. For these reasons, we decided to use standard gestures [65] and map those gestures to our features (in Figure 5.3). Standard iOS gestures include: tap, drag, flick, swipe, double tap, pinch, touch and hold, shake, and rotate. We made sure that our design is consistent with existing design practices and mental models that users are habituated with.

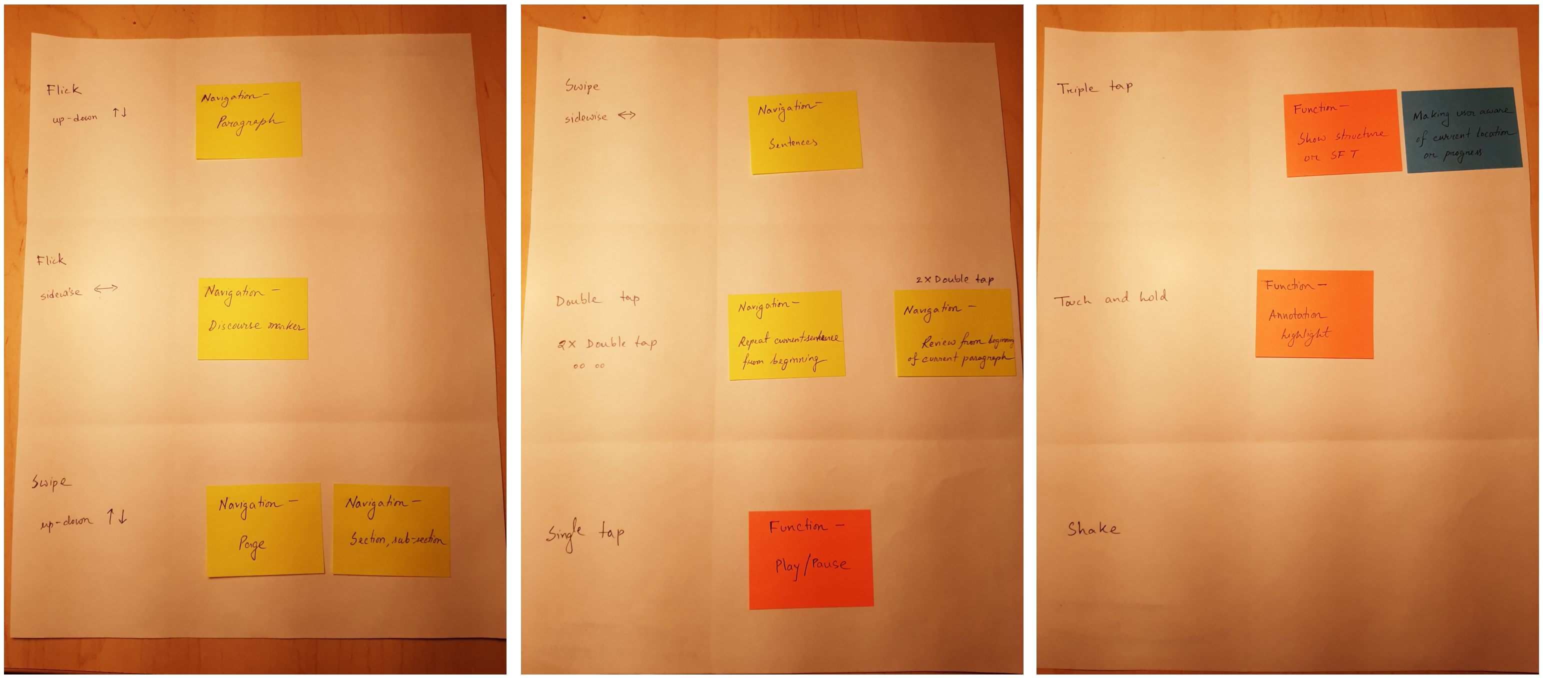
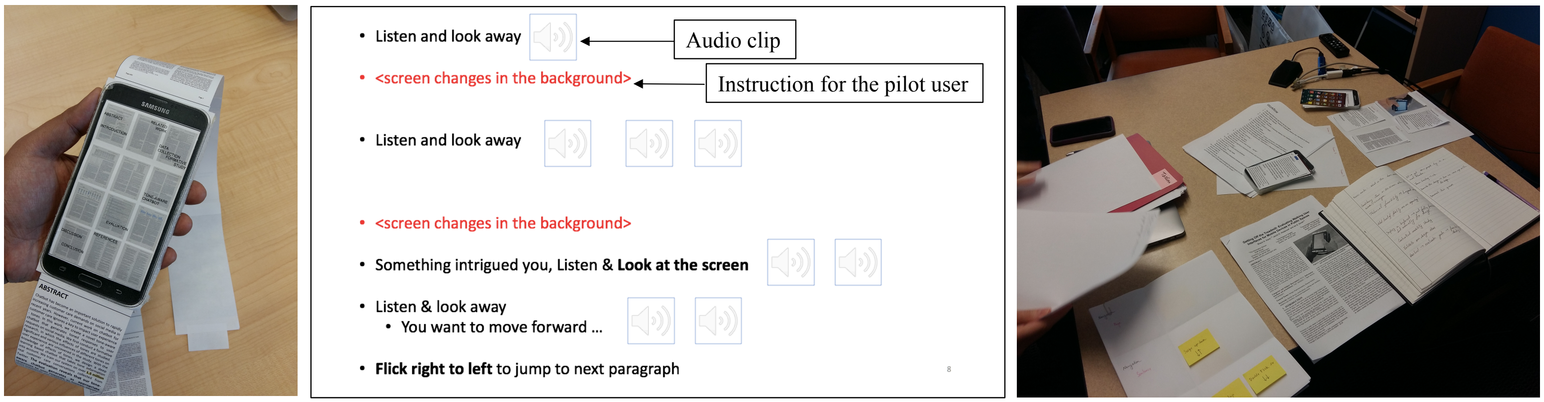


Figure 5.3: Mapping standard gestures to features.

5.2.2 Design and Prototyping Method Refinements

We iterated over our initial design and gathered feedback from pilot users. Based on the feedback received, we refined our design and simulated eyes-reduced skimming using a PowerPoint slide show.

Observations from our initial round of design indicated that users' experience with our prototyping method was too slow. This happened because in our first prototype (Figure 5.2), the interface stages were managed using paper screens. Pilot users would take off the current paper screen and put on next screen from a deck of paper screens. Another problem was that the users would mistakenly put the current screen on top of the next paper screen deck. This created confusion and users would lose valuable context while using the paper prototype. To solve this problem, we chained paper screens one after another in the order of their appearance (Figure 5.4 (a)). Users would pull out from the top to bring the next paper screen to view.



(a) (b) (c)

Figure 5.4: (a) Refined paper prototype - paper screens chained chronologically in the order of appearance. (b) PowerPoint slide with embedded audio files and instructions for both pilot user (in red ink) and the facilitator (in black ink). (c) Three-way setup for two pilot users and one facilitator in a session.

Simulating voice narration

To simulate voice narration in auditory skimming while users interacted with the paper prototype, we embedded audio clips in a PowerPoint slide show (Figure 5.4 (b)). The audio clips were extracted from different parts of a selected research article, such as the first few sentences of a paragraph, section or sub-section titles, sentences that referred to a figure or table, and captions of a figure or a table. We also outlined instructions on how to simulate voice narration in the slides.

To experience skimming using our paper prototype and simulated voice narration, we used a wizard-of-oz technique. Two members of our research team acted as pilot users and the thesis author acted as the facilitator. During this three-way session (in Figure 5.4 (c)), the facilitator would walk through the PowerPoint slides. Instructions in black ink were for the facilitator to read-out-loud to the pilot users. Instructions in red ink were for the pilot users, mostly asking to change the paper screens. Audio clips were embedded with these instructions, such that they would play automatically after some time.

Challenges of the low-fidelity prototype

Our main pain point while prototyping was synchronizing the audio with the physical navigation and spatial layout. Consider the last instruction from Figure 5.4 (a). The user had to flick from right to left and slide the paper screens to experience the visual change in output, while listening to the audio at the same time. Listening to the facilitator, while also maintaining an awareness of listening to the audio, seemed overwhelming to the pilot users. Instruction sequencing was also challenging for the facilitator too. Sometimes, audio clips would start and overlap with the facilitator's instruction, which created more confusion. To properly evaluate whether users can listen to a given text in practice, we had to build an interactive medium-fidelity prototype.

5.2.3 Building a Usable Prototype

We moved onto implementing an interactive medium-fidelity system that dynamically renders narrations and visual interfaces in response to user interactions on-the-fly. Figures 1.1 and 1.2 show snapshots of Skimmer.

Skimmer was implemented as a Javascript-based web app running on a 6.4-inch mobile phone. We built Skimmer to support HTML-based e-publication format as a source document type (e.g., an HTML version of a paper in the latest ACM CHI Proceedings). A couple of reasons encouraged us to move forward with HTML-based source documents. First, from our need finding study in Chapter 3, we learned that although participants preferred original layout over plain text layout, there are some benefits of plain text layout that can positively influence eyes-reduced skimming. An HTML-based document is browse-able in a single page by scrolling, rather than paginated. Second, as the e-pub scene in the research community is moving from static PDF to such flexible document types for enhancing accessibility and device-compatibility, we consider using the HTML version of a document to be a future-facing decision.

Skimmer's development consisted of the following high-level modules:

- *Audio module.* We used HTML5's audio library to build a simple audio module. HTML5 Audio defines some DOM methods, properties, and events that we used to in our audio module. For example, we used the basic *play()* and *pause()* to play and stop the audio. We used *playbackRate* property to set speech rate. The two most useful events are *ontimeupdate* and *onended*, because they played a crucial role to refresh and update the display module and signal the end of an audio clip respectively.

- *Data module.* For each text unit, Skimmer has an audio file and a JSON file timestamped at the word-level. We define text units as the section titles, a paragraph, or caption of a table or figure. We generated audio files by pre-processing text data manually. One example of text pre-processing is given in Chapter 5.1.2. We passed audio files to a speech recognition engine, which generated JSON files with word-level timestamps.

- *Display module.* In the display module, we highlighted the sentence and the word currently narrated. The display module is fired at each *ontimeupdate*. The word at that time instance is highlighted, including the sentence containing that word.

- *Touch gesture module.* Our decision to use HTML-based e-publication format implied that the application would eventually run on a web browser. The default web browser scrolling provides a rich scroll-based user experience. We wanted to have the same experience in our application. However, there are also some other default features that we did not want in our application. For example, double tap and press and hold would highlight text for selection in a web browser. Double tap may also cause zoom-in and out in some cases. We wanted to customize the result of gesture interaction, for example, tap on either side of the screen to navigate sentences, or flick left to right to change view. Therefore, our needs were special in that simply calling *preventdefault()* function to prevent the browser's default behavior and customize them based on our needs was not a straight-forward solution. We tried and tested several Javascript-based gesture libraries - such as ZingTouch, Hammer.JS, interact.js, and DeepTissue.js. After much deliberation and testing, we used ZingTouch to handle drag, swipe, and scroll based gestures, and Hammer.JS to handle tap based gestures, such as single tap, double tap, and press and hold. We included both of them because calling *preventdefault()* would break Hammer.JS and ZingTouch would not stop bubbling tap events.

In summary, we designed a usable prototype called Skimmer through multiple rounds of designs, implementations, and pilot evaluations.

Chapter 6

Preliminary Evaluation of Skimmer

We conducted a preliminary evaluation of Skimmer. It was a structured observation on a public bus where participants experienced both Skimmer and VDR (also used in our need finding study in Chapter 3). We used VDR as a reference point to ground participants' audio skimming experience, not with an expectation for it to be evaluated "head to head" (as VDR was not designed for auditory skimming). Our primary evaluation goal was to assess qualitatively how the participants experienced the unique features in Skimmer. All user study materials for this study can be found in Appendix B.

6.1 Method

Six graduate students (3 male, 3 female) from various disciplines took part (none from our need finding study in Chapter 3.). For the task, we asked participants to skim two documents in "eyes-reduced" manner using audio, one document with each application. The motivation, context, and protocol were similar to the previous study. As reading materials, we chose two articles from CHI 2019, both related to social media (hereafter referred to as A1 [66] and A2 [67]). Both articles are isomorphic in style, structure, and difficulty (on the Flesch-Kinkaid Reading Ease scale [55]). Each participant experienced both articles; orders of both articles

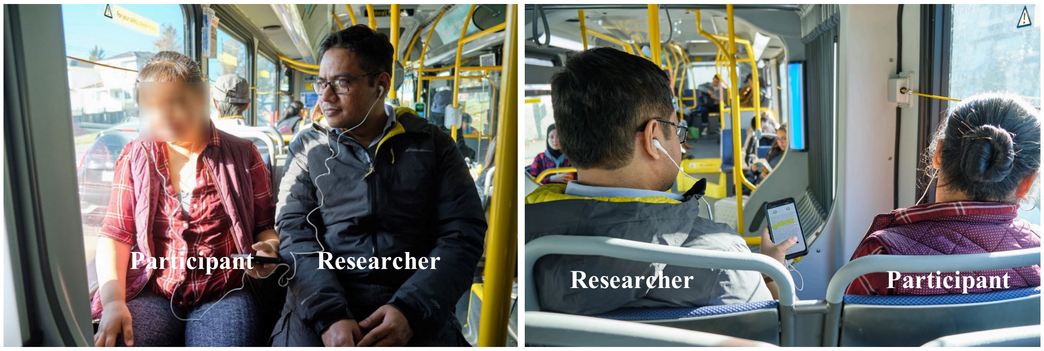


Figure 6.1: Field study to evaluate Skimmer. The study was conducted in public buses. A researcher accompanied a participant and conducted the study in a bus.

and apps was counterbalanced. Times allotted for skimming A1 and A2 was 9 and 8 minutes respective, due to differences in length (10,275 vs. 7,745 words). We followed the process for determining time to skim in Chapter 3. The procedure was similar to the previous study, except run on the bus. For each app, we first demonstrated it to the participants and let them play with it to become comfortable. The thesis author accompanied the participants in the bus, observed them completing their tasks and took notes (in Figure 6.1). Both Skimmer and VDR were deployed in a Samsung A50 Android device. We used a two-way audio splitter so that both the participant and the author could listen to the audio output from the app individually (in Figure 6.1). After each app+article, the participant and the author got off the bus. The participant answered a small number of questions and then sat for an interview (audio-recorded).

6.2 Findings

All 6 participants completed the tasks with both apps, in that they skim-read the articles. Three participants offered without prompting that they typically experience motion sickness.

Skimmer can be used eyes-reduced, VDR requires near constant visual attention. It was very clear in our observation that participants were able to largely not look at the screen using Skimmer. There were intermittent episodes where participants glanced at the screen: (1) in the Overview page to section jump, and (2) in the Article page to see interesting keywords (e.g., "Trump"), reacting to (unexpected) earcons (e.g., appearance of new section after next paragraph gesture), and responding to the haptic nudge when a figure or table was referred in the text. P2, P4, and P5 looked at the screen in the Results section, skipping by paragraphs, often stopping for a brief moment "to find interesting results quickly". The comparison to VDR was black and white: participants spent the majority of their time scrolling *"all the way down to find Conclusion. I don't have any other way of doing that other than looking at the screen. But it was really easy with this application (Skimmer)."* (P2)

Users appreciate Overview the most. Overview was called out as the most useful feature in Skimmer: "key difference between the two apps." Participants frequently navigated to the Overview page by swiping to the right (except P1 tapping Overview at the top one time), scrolling the list of sections and sub-sections and looking at the screen intermittently to find a section and then swiping to the left, eyes-free to complete a jump to that section. At one moment, P3 and P4 were tapping instead of swiping to the left because they "forgot how to go back", which also prompted them to look at the screen. In contrast, such section jumps prompted visual attention, heavy and frequent scrolling in VDR.

Auditory feedback and haptic affordance help to re-focus. While the intended purpose of the auditory and haptic cues was to subtly reinforce an interaction (e.g., next sentence) or signal an affordance (e.g., image available to see), they also helped participants "to situate themselves and come back" when they lost focus. P5 said, *"I would say, the earcon for Section was really helpful. I was looking outside and maybe lost focus, that earcon helped me situate myself again and come back."* However, some participants did not notice the earcons at all. This could mean that our choice of earcons in our design was successful, in that participants did not feel that these earcons were rather annoying them or contributed to losing focus from listening in any way.

Individual differences in navigation preferences are supported. P2 jumped from Conclusion to the last sentence of Abstract by using a combination of section, paragraph, and sentence jumps as “shortcuts”. P1, P3, and P6 used sentence jumps frequently because “it was faster”, while P4 and P5 used drag for sentence jumps frequently and flick for paragraph jumps occasionally because “scrolling felt natural”.

Skimming numbers remains challenging for those who tried it. With the exception of P5 and P6, other participants did not consider numbers important in their skimming. P5 looked down upon numbers in the Results section and visually read with both VDR and Skimmer because *“it was difficult to follow numbers, and defeated the purpose of audio skimming.”* Both P2 and P5 resonated that *“numbers should come with context, for example, 1500 participants”* instead of just “1500”.

Users appreciate narration quality and disparate voices. Participants were impressed by the quality of narration in Skimmer, mentioning that it was *“soft and comfortable to listen to, unlike the stiff voice in the other (VDR)”*. In contrast, P5 mentioned that she liked VDR’s narration because it was *“natural, research papers are dry, you don’t need emotion here, a robotic voice is also okay.”* Skimmer’s use of disparate voices was “unique” but little noticed.

Discourse markers are useful, but need more exposure. Participants described the discourse marker jump as a “great tool” and “helpful”. Only P3 and P5 used it in the actual study. Regardless, all of them experienced it during training too, and mentioned that they “needed more trust.” P1 said, *“I may need to experience it just by myself alone before I know what really is in there”*. In contrast, we observed similar patterns of “zoom-in, spot discourse markers” behavior from need finding study while using VDR.

Figures/Tables are mostly ignored, but appreciated the idea of a haptic nudge. Participants thought that haptic nudge was a *“a great way of reminding me that something happening here.”* (P3) Only P1 and P3 used this opt-in feature to view a figure and a table respectively, while the rest of the participants ignored, citing “figures/tables not important in skimming”. P3 said, *“I don’t do audio reading* *because I am not good at keeping attention. But such vibrations are really useful."* In contrast, while using VDR, P1 and P3 momentarily stopped at a figure/table while scrolling visually, zoomed-in and out, and then moved on.

Participants appreciated Skimmer's design concept. For example, we heard - "lots of reading for a PhD student, not a terrible idea to get gist for some... super helpful in skimming qualitative papers"(P5), "Pocket for commuting"(P6), and even the suggestion that it could be used for parallel visual and auditory reading "read-along for listening and following words"(P2). P6's experience highlights one of our key research motivations: she felt nauseated while using VDR and had to stop momentarily, and by contrast she did not report feeling nauseated while using Skimmer and was able to complete the task eyes-reduced.

Participants performed comparably on the reading comprehension test. The statistic show no significant differences on comprehension. This is not surprising given that with VDR they predominantly used their uses to skim, whereas with Skimmer they predominantly listened.

Chapter 7

Discussion

We reflect on our key findings and their implications for design and our design process.

We have made concrete progress towards eyes-reduced skimming. Our goal for this research was to design an interface that would allow people to skim documents in an eyes-reduced way. Given the complexity of skimming as a form of reading [40, 3] that frequently mixes between non-linear reading, and glancing, supporting it via the audio channel is a non-trivial design objective. Especially for an on-the-go setting, where distractions and situational impairment factors are prevalent, it poses more challenges. In our preliminary evaluation *on the bus*, we could only ask participants to try not to use their eyes, but we could not force them to not look. We were pleasantly surprised at the extent to which participants did in fact not use their eyes while using Skimmer, which was not possible with the traditional read-aloud application. Further, our participants were positive about the overall design concept of Skimmer, with 5 of them indicating that they would use it to skim documents. This suggests that we have indeed made positive progress towards our design goal, that the Skimmer design is on the right track.

Skimmer is a design snapshot of an artefact in its evolution. Even though results from our preliminary evaluation were promising, there are opportunities to improve usability and address the gaps in our design for supporting eyes-reduced interaction. For example, TTS needs to go beyond simply narrating the given text with verbatim speech. There is room for improvement not only in what they say (content) but also how they say it (paralingual cues, including voice modulation, varying pitch, tone, and varying speech rate) for better comprehension. We only explored some of those dimensions in our design, such as numbers and lists. Further, new elements such as a means to navigate to and from metadata such as headers, footers, and references, will need to be integrated such that they do not make the interaction overly complex. More substantive design steps, towards fully eyes-free design, are described further below.

Eyes-reduced design is a new concept. Our work has introduced the concept of eyes-reduced design – it is intended for designs that can be used primarily eyes-free, but, importantly, do not preclude the user from using their eyes, when they want. This can be differentiated from design that is created with the intention of being used with one's eyes, but also happens to support some degree of eyes-free interaction. For example, a person can use YouTube on a mobile device while on the go. They can select a video while looking and then, to some extent, look away while walking, toggling the pause/play to listen to the video by tapping on the screen. This is challenging because the region that accepts the tap to pause/play is small, specified by a visual icon. Thus, while it can be used-eyes free for some interactions, it is difficult to do so, and was not seemingly designed with that intention.

Fully eyes-free skimming is next step. While Skimmer was designed with the intention that it be usable largely eyes-free, we have not yet tested its eyes-free efficacy. Fully eyes-free skimming is the next main target of our research. Our goal from the outset of this research was that Skimmer support both the needs of people experiencing situational impairments as well as those who have visual impairments. We made the choice to start our design process with targeting situational impairments, as we reckoned it may be the more difficult design problem – people in situational impairments still have an option to use their eyes (hence eyes-reduced) when they deem it necessary. So the design would need to support fluidly moving back and forth between primarily using one's eyes and one's audition. Now that we have made progress on the eyes-reduced front, we will turn our attention to assessing the extent to which Skimmer can be used *entirely eyes-free*. Some extensions are immediately obvious, such as reading out figure descriptions (alt text), but then our work will continue with individuals with visual impairments, including a systematic needs analysis. Evaluation will undoubtedly reveal that some of Skimmer's design elements need adjustment to support that population. Upon iterating on the design, we will loop back to evaluating with people in situational impairments so as to ensure that our design is maximally supporting both eyes-reduced and eyes-free interaction.

Preset reading order and eyes-reduced annotation capability are also next steps. We chose not to implement DG-10 and DG-11 in our first run of implementation. This is because we wanted to evaluate our design first, before implementing features that bring added values to the very primary and basic needs. Some participants from our need finding study in Chapter 3 mentioned the need for a preset reading order and annotation capability while skimming.

Preset reading order means automatically reading through an order set by the user, such as Abstract, Introduction, Conclusion, and Discussion. This is an advanced feature. Users often have a preconceived idea of what to read and in which order to read when the document is familiar [43]. Our need finding study in Chapter 3 also showed evidence of this behavior among our participants. There are two ways this feature can be implemented. First, the app can learn user's reading behaviors and patterns and predict the future reading pattern for a similar class of document. This is similar to recommendation systems (ads, search engines) that learn from our past history. Second, the app can have a settings panel where users can select the relative ordering of chapters or sections. This is similar to creating a playlist and listening in the order the user has selected. We anticipate some challenges while implementing this feature. Preset reading order may not be applicable to some documents, such as online news articles, which do not have a systematic structure. Structural outline varies among scientific documents too, which requires further consideration (e.g., ACM CHI papers and Nature papers have different structural outline).

One of the fundamental goals of skimming is assessing whether the reading material warrants further attention from the reader [9]. This brings annotation capability in skimming context, because people often highlight or comment (putting stars or squiggle) texts within a document to indicate content that is important, or content that they need to revisit later [62]. Such kinds of annotation are called within-text telegraphic annotation. In our study, we only evaluated a single pass skimming episode to evaluate our design. Hence, we did not design for DG-11. In our future work, we will explore ways and means of eyes-reduced annotation creation and consumption. Design ideas can include eyes-free annotation creation, such as press and hold while narration is ongoing, indicating to the system to highlight the underlying text. Another possibility can be speech recorded annotation. This can be anchored to the text currently narrated, without requiring the user to look at the screen. For consumption, there should be a way to navigate annotations. Further, one possibility is also to have a cloud-based system, where the user will be able to skim and annotate while on-the-go, for example, and then perhaps if she later decides to revisit the content again when she is back to her desk, she can access the same annotated content when making a detailed pass on it. We anticipate some challenges here too. For example, speech recorded annotation may be uncomfortable and undesirable in public places. Further, it is an interesting problem how to inform the user about an annotation in the current content. Should the system vibrate or make an auditory feedback? This may work if annotations are sparse and limited in number. How should the system let the user know if the current sentence is highlighted or underlined?

Different environments/contexts can pose different kinds of situational impairment. We only explored a context where people could not fully use their *visual* attention because of potential motion sickness. However, different environments/contexts result in situational impairments that vary in degree. For example, one riding a commuting bus can occasionally glance their mobile, but it will be nearly impossible for one driving a vehicle to use their eyes even momentarily.

Also, a specific impairment context may pose challenges in using interaction modalities other than eyes. Zeleznik et al., documented the ‘sandwich’ problem when a user is holding an object in one hand, which impairs one’s manual interaction capability [68]. We wonder about other types of situational impairment that variably affect different modalities. For example, what about when people can see, but cannot touch the device (e.g., reading recipes when your hands are oily or wet)? Interfaces that address specific kinds of situational impairments is a promising direction. Beyond eyes-reduced. What about touch-reduced?

Our design guidelines can inspire future design directions. We derived our design guidelines in Chapter 4 from multiple research fronts. Our intention was to explore relevant research areas and put up a robust design guideline which can motivate designers to come up with new design ideas, perhaps better than Skimmer.

We can think of some design directions which can positively influence user experience of Skimmer. First, similar to how Ahmed et al., [48, 49, 50] leverage semantic summary of text using Natural Language Processing (NLP) techniques, Skimmer can also leverage semantic summary of text and provide a summary snapshot of the text besides skimming. Second, our eyes-reduced design currently acknowledges visual attention only in cases where a figure or table is referred. What if the user wants to read-along? Using the front camera and eye-tracking technique, the system can enlarge the text where user’s eyes are focused and facilitate read-along when happening.

Chapter 8

Limitations

There are limitations to our work. We focused on structured, academic/professional documents. It remains to be seen how well Skimmer will work with documents that have a lot of visuals or ones with many links that require cross-document navigation, thus beyond non-linear. It is possible that a different design would be needed altogether. Further, our evaluation of Skimmer only involved 6 graduate students, with a single exposure. A larger sample from a more varied population, together with a longer exposure time to Skimmer, will be needed to more deeply assess the design elements.

Chapter 9

Conclusion

In this work we tackled the design challenge of supporting eyes-reduced skimming of structured documents. Based on our need finding study, we generated design guidelines for eyes-reduced skimming, iterated extensively on the design based on the guidelines, and then implemented Skimmer, a prototype app. Through a preliminary evaluation, we learned that people could use Skimmer eyes-reduced while riding the bus, only looking down at the app infrequently, as they navigated a CHI paper largely eyes-free, listening to the content.

We are conducting our work within the framework of inclusive design [69], with planned next steps to involve people with visual impairments so that we can push more strongly on eyes-free interaction. This will lead us one step closer to achieving a system that people who are experiencing impairments (either situationally or more permanently) can use to skim documents.

Bibliography

[1] R. Sandler, "How much data do we create every day? the mind-blowing stats everyone should read," June 2018, retrieved September 19, 2019 from https://www.businessinsider.com/mozilla-scout-voice-controlled-browser-2018-6.

[2] R. I. L. Inc., "Pocket," 2019, retrieved September 19, 2019 from https://getpocket.com/.

[3] G. B. Duggan and S. J. Payne, "Skim reading by satisficing: Evidence from eye tracking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 1141–1150. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979114

[4] A. Thayer, C. P. Lee, L. H. Hwang, H. Sales, P. Sen, and N. Dalal, "The imposition and superimposition of digital reading technology: The academic potential of e-readers," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2917–2926. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979375

[5] B. Arons, "Speechskimmer: Interactively skimming recorded speech," in *Proceedings of the 6th Annual ACM Symposium on User Interface Software*

and Technology, ser. UIST '93. New York, NY, USA: ACM, 1993, pp. 187–196. [Online]. Available: http://doi.acm.org/10.1145/168642.168661

[6] S. Bly and C. C. Marshall, "Turning the page on navigation," in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*. IEEE, 2005, pp. 225–234.

[7] M. Wolf, "Skim reading is the new normal. the effect on society is profound," August 2018, retrieved September 18, 2019 from https://www.theguardian.com/commentisfree/2018/aug/25/skim-reading-new-normal-maryanne-wolf.

[8] B. Merr, "How much data do we create every day? the mind-blowing stats everyone should read," May 2018, retrieved September 18, 2019 from https://bit.ly/2krBBrv.

[9] M. J. Adler and C. Van Doren, *How to read a book: The classic guide to intelligent reading*. Simon and Schuster, 2014.

[10] J. M. Hailpern and B. A. Huberman, "Odin: Contextual document opinions on the go," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 1525–1534. [Online]. Available: http://doi.acm.org/10.1145/2556288.2556959

[11] B. Yi, X. Cao, M. Fjeld, and S. Zhao, "Exploring user motivations for eyes-free interaction on mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 2789–2792. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208678

[12] C. Ingraham, "The astonishing human potential wasted on commutes," February 2016, retrieved September 18, 2019 from

https://www.washingtonpost.com/news/wonk/wp/2016/02/25/how-much-of-your-life-youre-wasting-on-your-commute/?noredirect=on.

[13] A. Sears, M. Lin, J. Jacko, and Y. Xiao, "When computers fade: Pervasive computing and situationally-induced impairments and disabilities," in *HCI international*, vol. 2, no. 03, 2003, pp. 1298–1302.

[14] J. O. Wobbrock, "Situationally-induced impairments and disabilities," in *Web Accessibility*. Springer, 2019, pp. 59–92.

[15] Z. Sarsenbayeva, N. van Berkel, C. Luo, V. Kostakos, and J. Goncalves, "Challenges of situational impairments during interaction with mobile devices," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ser. OZCHI '17. New York, NY, USA: ACM, 2017, pp. 477–481. [Online]. Available: http://doi.acm.org/10.1145/3152771.3156161

[16] M. Lin, R. Goldman, K. J. Price, A. Sears, and J. Jacko, "How do people tap when walking? an empirical investigation of nomadic data entry," *International journal of human-computer studies*, vol. 65, no. 9, pp. 759–769, 2007.

[17] A. Ng, S. A. Brewster, and J. H. Williamson, "Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 1981–1990.

[18] M. E. Mott and J. O. Wobbrock, "Cluster touch: Improving touch accuracy on smartphones for people with motor and situational impairments," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019, pp. 27:1–27:14. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300257

[19] S. K. Kane, J. O. Wobbrock, and I. E. Smith, "Getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces,"

in *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '08. New York, NY, USA: ACM, 2008, pp. 109–118. [Online]. Available: http://doi.acm.org/10.1145/1409240.1409253

[20] J. D. Hincapié-Ramos and P. Irani, "Crashalert: Enhancing peripheral alertness for eyes-busy mobile interaction while walking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 3385–3388. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466463

[21] A. Bragdon, E. Nelson, Y. Li, and K. Hinckley, "Experimental analysis of touch-screen gesture designs in mobile environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 403–412. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979000

[22] D. Ghosh, P. S. Foong, S. Zhao, D. Chen, and M. Fjeld, "Editalk: Towards designing eyes-free interactions for mobile word processing," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 403:1–403:10. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173977

[23] B. Schildbach and E. Rukzio, "Investigating selection and reading performance on a mobile phone while walking," in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '10. New York, NY, USA: ACM, 2010, pp. 93–102. [Online]. Available: http://doi.acm.org/10.1145/1851600.1851619

[24] K. Vadas, N. Patel, K. Lyons, T. Starner, and J. Jacko, "Reading on-the-go: A comparison of audio and hand-held displays," in *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and*

Services, ser. MobileHCI '06. New York, NY, USA: ACM, 2006, pp. 219–226. [Online]. Available: http://doi.acm.org/10.1145/1152215.1152262

[25] S. Carter and L. Denoue, "Seereader: An (almost) eyes-free mobile rich document viewer," *arXiv preprint arXiv:0909.2185*, 2009.

[26] C.-H. Yu and R. C. Miller, "Enhancing mobile browsing and reading," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '11. New York, NY, USA: ACM, 2011, pp. 1783–1788. [Online]. Available: http://doi.acm.org/10.1145/1979742.1979845

[27] O. Metatla, N. Bryan-Kinns, and T. Stockman, ""i hear you": Understanding awareness information exchange in an audio-only workspace," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 546:1–546:13. [Online]. Available: http://doi.acm.org/10.1145/3173574.3174120

[28] Y. Vazquez-Alvarez and S. A. Brewster, "Eyes-free multitasking: The effect of cognitive load on mobile spatial audio interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2173–2176. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979258

[29] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, "Multimodal 'eyes-free' interaction techniques for wearable devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03. New York, NY, USA: ACM, 2003, pp. 473–480. [Online]. Available: http://doi.acm.org/10.1145/642611.642694

[30] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch, "Earpod: Eyes-free menu selection using touch input and reactive audio feedback," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. New York, NY, USA: ACM, 2007,

pp. 1395–1404. [Online]. Available: http://doi.acm.org/10.1145/1240624.1240836

[31] K. A. Li, P. Baudisch, and K. Hinckley, "Blindsight: Eyes-free access to mobile phones," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1389–1398. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357273

[32] D. Ashbrook, P. Baudisch, and S. White, "Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2043–2046. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979238

[33] J. Pasquero, S. J. Stobbe, and N. Stonehouse, "A haptic wristwatch for eyes-free interactions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 3257–3266. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979425

[34] M. Pielot, A. Kazakova, T. Hesselmann, W. Heuten, and S. Boll, "Pocketmenu: Non-visual menus for touch screen devices," in *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, ser. MobileHCI '12. New York, NY, USA: ACM, 2012, pp. 327–330. [Online]. Available: http://doi.acm.org/10.1145/2371574.2371624

[35] S. Zhu, J. Zheng, S. Zhai, and X. Bi, "I'sfree: Eyes-free gesture typing via a touch-enabled remote control," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: ACM, 2019, pp. 448:1–448:12. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300678

[36] S. Holland, D. R. Morse, and H. Gedenryd, "Audiogps: Spatial audio navigation with a minimal attention interface," *Personal and Ubiquitous computing*, vol. 6, no. 4, pp. 253–259, 2002.

[37] D. Bragg, C. Bennett, K. Reinecke, and R. Ladner, "A large inclusive study of human listening rates," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 444:1–444:12. [Online]. Available: http://doi.acm.org/10.1145/3173574.3174018

[38] M. H. Cohen, M. H. Cohen, J. P. Giangola, and J. Balogh, *Voice user interface design*. Addison-Wesley Professional, 2004.

[39] M. E. Masson, "Cognitive processes in skimming stories." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 8, no. 5, p. 400, 1982.

[40] G. B. Duggan and S. J. Payne, "Text skimming: The process and effectiveness of foraging through text under time pressure." *Journal of Experimental Psychology: Applied*, vol. 15, no. 3, p. 228, 2009. [Online]. Available: http://dx.doi.org/10.1037/a0016995

[41] C. C. Marshall and C. Ruotolo, "Reading-in-the-small: A study of reading on small form factor devices," in *Proceedings of the 2Nd ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '02. New York, NY, USA: ACM, 2002, pp. 56–64. [Online]. Available: http://doi.acm.org/10.1145/544220.544230

[42] T. Hillesund, "Digital reading spaces: How expert readers handle books, the web and electronic paper," 2010.

[43] J. Franze, K. Marriott, and M. Wybrow, "What academics want when reading digitally," in *Proceedings of the 2014 ACM Symposium on Document*

Engineering, ser. DocEng '14. New York, NY, USA: ACM, 2014, pp. 199–202. [Online]. Available: http://doi.acm.org/10.1145/2644866.2644894

[44] D. Kim, G. Gweon, and G. Lee, "Sats: Structure-aware touch-based scrolling," *ETRI Journal*, vol. 38, no. 6, pp. 1104–1113, 2016.

[45] B. Lee, O. Savisaari, and A. Oulasvirta, "Spotlights: Attention-optimized highlights for skim reading," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 5203–5214. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858299

[46] S. Whittaker, J. Hirschberg, J. Choi, D. Hindle, F. Pereira, and A. Singhal, "Scan: Designing and evaluating user interfaces to support retrieval from speech archives," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 26–33.

[47] S. Vemuri, P. DeCamp, W. Bender, and C. Schmandt, "Improving speech playback using time-compression and speech recognition," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 295–302. [Online]. Available: http://doi.acm.org/10.1145/985692.985730

[48] F. Ahmed, Y. Borodin, Y. Puzis, and I. V. Ramakrishnan, "Why read if you can skim: Towards enabling faster screen reading," in *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, ser. W4A '12. New York, NY, USA: ACM, 2012, pp. 39:1–39:10. [Online]. Available: http://doi.acm.org/10.1145/2207016.2207052

[49] F. Ahmed, Y. Borodin, A. Soviak, M. Islam, I. Ramakrishnan, and T. Hedgpeth, "Accessible skimming: Faster screen reading of web pages," in *Proceedings of the 25th Annual ACM Symposium on User Interface Software*

and Technology, ser. UIST '12. New York, NY, USA: ACM, 2012, pp. 367–378. [Online]. Available: http://doi.acm.org/10.1145/2380116.2380164

[50] F. Ahmed, A. Soviak, Y. Borodin, and I. Ramakrishnan, "Non-visual skimming on touch-screen devices," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, ser. IUI '13. New York, NY, USA: ACM, 2013, pp. 435–444. [Online]. Available: http://doi.acm.org/10.1145/2449396.2449452

[51] M. Büchi, N. Just, and M. Latzer, "Caring is not enough: the importance of internet skills for online privacy protection," *Information, Communication & Society*, vol. 20, no. 8, pp. 1261–1278, 2017.

[52] A. Sulaiman, N. I. Jaafar, and A. Tamjidyamcholo, "Influence of personality traits on facebook engagement and their effects on socialization behavior and satisfaction with university life," *Information, Communication & Society*, vol. 21, no. 10, pp. 1506–1521, 2018.

[53] M. Anderson and J. Jiang, "Teens, social media technology 2018," Pew Research Center, May 2018, retrieved September 18, 2019 from https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/.

[54] A. Smith and M. Duggan, "Crossing the line: What counts as online harassment?" Pew Research Center, January 2018, retrieved September 18, 2019 from https://www.pewinternet.org/2018/01/04/crossing-the-line-what-counts-as-online-harassment/.

[55] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," 1975.

[56] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[57] K. Charmaz and L. L. Belgrave, "Grounded theory," *The Blackwell encyclopedia of sociology*, 2007.

[58] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 143–146. [Online]. Available: http://doi.acm.org/10.1145/1978942.1978963

[59] B. Shneiderman, "The limits of speech recognition," *Communications of the ACM*, vol. 43, no. 9, pp. 63–65, 2000.

[60] D. Goldberg, "The voice actor's performance guidebook," 2019, retrieved September 19, 2019 from https://www.edgestudio.com/voice-over-performance-guidebook.

[61] K. Hyland and P. Tse, "Metadiscourse in academic writing: A reappraisal," *Applied linguistics*, vol. 25, no. 2, pp. 156–177, 2004.

[62] C. C. Marshall, "Annotation: from paper books to the digital library," in *Proceedings of the second ACM international conference on Digital libraries*. ACM, 1997, pp. 131–140.

[63] H. Seifi, K. Zhang, and K. E. MacLean, "Vibviz: Organizing, visualizing and navigating vibration libraries," in *2015 IEEE World Haptics Conference (WHC)*. IEEE, 2015, pp. 254–259.

[64] A. Cockburn, C. Gutwin, and J. Alexander, "Faster document navigation with space-filling thumbnails," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: ACM, 2006, pp. 1–10. [Online]. Available: http://doi.acm.org/10.1145/1124772.1124774

[65] A. Inc., “Human interface guidelines,” retrieved September 18, 2019 from https://developer.apple.com/design/human-interface-guidelines/ios/user-interaction/gestures/#standard-gestures.

[66] H. Habib, N. Shah, and R. Vaish, “Impact of contextual factors on snapchat public sharing,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: ACM, 2019, pp. 26:1–26:13. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300256

[67] F. Bentley, K. Quehl, J. Wirfs-Brock, and M. Bica, “Understanding online news behaviors,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: ACM, 2019, pp. 590:1–590:11. [Online]. Available: http://doi.acm.org/10.1145/3290605.3300820

[68] R. Zeleznik, A. Bragdon, F. Adeputra, and H.-S. Ko, “Hands-on math: A page-based multi-touch and pen desktop for technical work and problem solving,” in *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’10. New York, NY, USA: ACM, 2010, pp. 17–26. [Online]. Available: http://doi.acm.org/10.1145/1866029.1866035

[69] P. J. Clarkson, R. Coleman, S. Keates, and C. Lebbon, *Inclusive design: Design for the whole population*. Springer Science & Business Media, 2013.

[70] J. Xu, S. Zhu, A. Balasubramanian, X. Bi, and R. Shilkrot, “Ultra-low-power mode for screenless mobile interaction,” in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’18. New York, NY, USA: ACM, 2018, pp. 557–568. [Online]. Available: http://doi.acm.org/10.1145/3242587.3242614

[71] D. Ghosh, P. S. Foong, S. Zhao, D. Chen, and M. Fjeld, "Editalk: Towards designing eyes-free interactions for mobile word processing," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 403:1–403:10. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173977

[72] A. Sears, M. Lin, J. Jacko, and Y. Xiao, "When computers fade: Pervasive computing and situationally-induced impairments and disabilities," in *HCI international*, vol. 2, no. 03, 2003, pp. 1298–1302.

[73] M. Goel, L. Findlater, and J. Wobbrock, "Walktype: Using accelerometer data to accomodate situational impairments in mobile touch screen text entry," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 2687–2696. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208662

[74] Z. Liu, "Reading behavior in the digital environment: Changes in reading behavior over the past ten years," *Journal of documentation*, vol. 61, no. 6, pp. 700–712, 2005.

[75] M. Viswanathan and M. Viswanathan, "Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (mos) scale," *Computer Speech & Language*, vol. 19, no. 1, pp. 55–83, 2005.

[76] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "Getting the right design and the design right," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: ACM, 2006, pp. 1243–1252. [Online]. Available: http://doi.acm.org/10.1145/1124772.1124960

Appendix A

Study Materials Used in Need Finding Study

In the following, we provide materials from our first study, the need finding study in Chapter 3. Specifically, we provide the following -

- Call For Participation
- Consent Form
- Study Protocol and Participant Training
- Questionnaires

A.1 Call For Participation

Department of Computer Science
University of British Columbia
Vancouver, BC, V6T 1Z4

Are you a grad student? Is it difficult for you to read research articles using a smartphone while commuting? Come help us as we try to make it easier for you!

Receive a $15 honorarium for this 1 hour study

Who are we?

We are researchers from the University of British Columbia: Dr. Joanna McGrenere, Dr. Dongwook Yoon, and Taslim Arefin Khan. We are conducting a study to better understand the needs and challenges of reading a document in situations where visual interaction with smartphone is challenging, e.g., commuting.

We are looking for participants who ...

- are graduate students
- can understand and speak English fluently
- are familiar with reading research articles published in conferences, journals, or online portals
- are 18 years or older

What is involved?

You will read few research articles using a smartphone application in a lab setting that will simulate bus commuting environment. After that, you will be interviewed about your experience of using that application. The study will be conducted at the ICICS Building, UBC, Vancouver.

Interested in Participating?

If you are interested in participating or would like more information, please contact Taslim at [redacted].

A.2 Consent Form for Participants

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

Consent Form

Non-visual document skimming on the go

Principal Investigator:
Joanna McGrenere, Professor, Department of Computer Science, University of British Columbia, [redacted].

Co-Principal Investigator:
Dongwook Yoon, Assistant Professor, Department of Computer Science, University of British Columbia, [redacted].
Taslim Arefin Khan, MSc student, Department of Computer Science, University of British Columbia, [redacted].

What the study is about:
The overall purpose of this research is to identify problems, needs, and requirement elicitations of audio-based reading (listening) using smartphones in situations where visual interaction with the smartphone is challenging.

What you will be asked to do: After you have read this document, I will respond to any questions or concerns that you may have. Once you have signed this consent form, you will be asked to:
- interact with a smartphone and read an audio-based reading material
- answer interview questions

This should take approximately 30 minutes to 1 hour and be completed in 1 session.

The session may also be photographed, video, and/or audio recorded. You have the option not to be photographed/video/audio recorded.

Inclusion criteria: We are looking for participants who:
- are graduate students
- can understand and speak English fluently
- are familiar with reading scholarly articles published in conferences or journals in their respective research fields
- are 18 years or older.

Risks and Benefits: There are no known psychological, physical, and social risks involved in the proposed research. Participants may learn how to benefit from non-visual skim reading while commuting to accomplish a reading task.

How the data collected will be used: Data collected (including any image, audio/video recordings) will be used for analysis. This analysis will help us extract requirements for a skim reading interface. The research findings from this data may also be used to publish scholarly articles in conferences or journals.

Compensation: You will receive $15 for his one hour study. The compensation will be provided prior to the study as cash only.

Confidentiality: The results of your participation will be reported without any reference to you specifically. All information that you provide will be stored in Canada. Only the researchers involved in this research will have access to the data.

Your data may be used but will remain anonymous in any reports, research papers, thesis documents, and presentations that result from this work. Your name will never be used, only a code associated with the data. Any identifiable element in the images, videos, or audios will be blurred or obscured to prevent identification. The only link between the code associated and your actual name will be this consent form and a printed spreadsheet, that will be stored in a locked cabinet, in a research lab with controlled access at UBC. The file linking your name with the associated code will be kept separate from the data.

Data Retention: All electronic files will be stored in an encrypted hard drive of a password protected laptop. Backups of the data will be stored on the UBC Computer Science secure department servers. Any handwritten notes and paper transcripts will be kept in a locked cabinet in the researchers' laboratory or the Principal Investigator's office with controlled access at UBC. Five years after the completion of the research, the electronic data will be deleted, and all physical media, audio/video records, and paper transcripts will be shredded by the Principal Investigator.

Taking part is voluntary: Your participation is voluntary. You are free to stop your participation at any point. Refusal to participate or withdrawal of your consent or discontinued participation in the study will not result in any penalty or loss of benefits or rights to which you might otherwise be entitled. The Principal Investigator may at his/her discretion remove you from the study for a number of reasons. In such an event, you will not suffer any penalty or loss of benefits or rights which you might otherwise be entitled. It will not affect your current or future relationship with the University of British Columbia.

Ethical Concerns and Participant Rights

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Please tick at least one of the following

- ❏ I consent to be audio recorded in this study
- ❏ I consent to be video recorded in this study
- ❏ I consent to be photographed in this study
- ❏ I do not consent to be photographed, audio, and video recorded in this study

I, ________________________________, have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in this study. However, I realize that my participation is voluntary and that I am free to withdraw at any time.

____

Participant's Signature   Date

A.3 Study Protocol and Participant Training

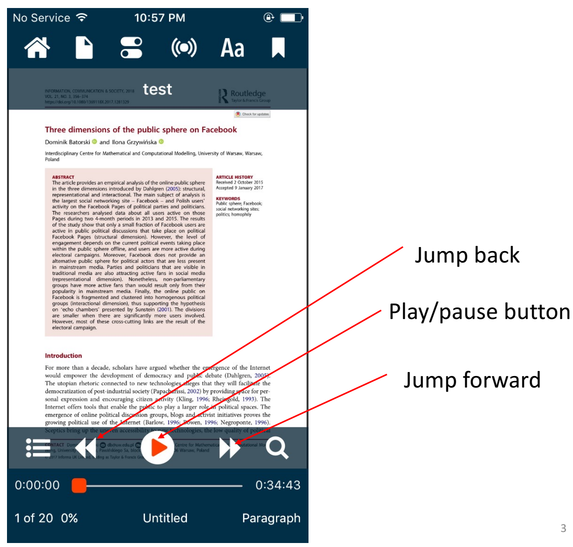
Speech Skimmer Study

Taslim Arefin Khan

1

Training with Voice Dream Reader App

Basic Controls

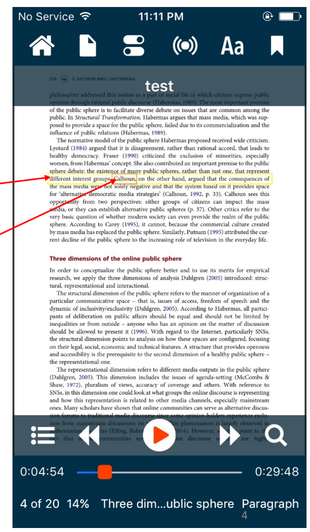


3

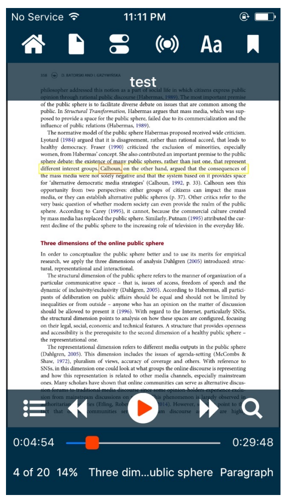
Following the Speech

Currently read line highlighted by a yellow rectangle

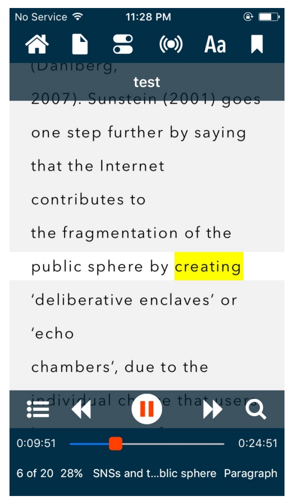
Currently read word highlighted by a red rectangle



Layouts



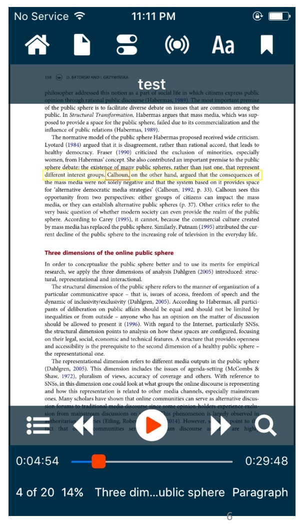
Original layout



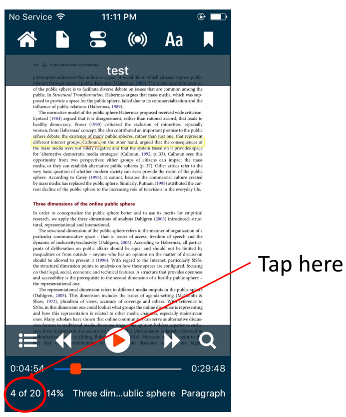
Plain text layout

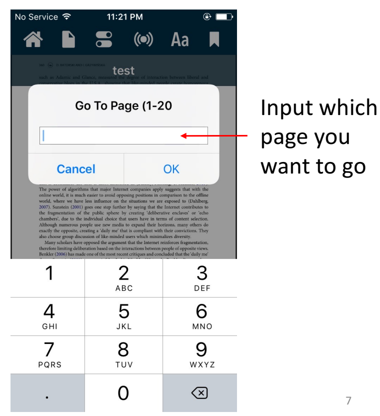
Read a Selected Line

Double tap on a line that you want to read



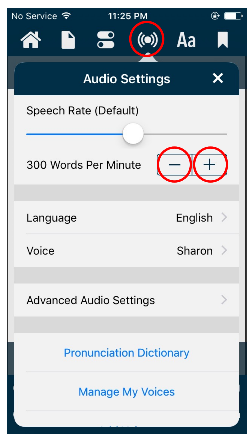
Jump to Another Page





7

Controlling Your Speed



Initially we will set speech rate to 250 WPM

You can adjust the speech rate to your comfort using the slider, or the + and - buttons

8

Your Task

You are a grad student. You are taking a grad course. For the next class, you have 4 reading assignments. You have to choose 2 articles which you will read in details. You want to participate in-class discussion where you have to reason why you have chosen those 2 articles.

You have a commute to school and just have enough time to read 2 articles after reaching school. You would like to use this commute time and skim all 4 articles to decide which 2 you are going to read later.

9

Your Objective

- You should skim each article to comprehend the following
  - Get a general idea or gist of the article
    - What is the article about? What problems are they working on?
    - What suggestions/takeaways have been discussed in the article?
  - Discover the structure of the material
    - How is the article structured? i.e., What is the outline in terms of sections?
    - Which parts/sections will help you most to understand the article when you read later?

How You Gonna Do It?

- You will perform each task in a standing posture
- After each task, you will answer a survey
- Your objective will be to finish each article under a time constraint
  - Minimize your visual attention to the screen as much as possible, else you will get motion sickness
  - You can look at the screen while putting a command, but look away while listening; if possible put the commands eyes-free

11

A.4 Questionnaires and Interview Questions

Non-visual Document Skimming On the Go Study

Demographic questions

- Which of following age group are you a part of?
    - 18 - 24
    - 25 - 34
    - 35 - 44
    - 45 - 54
    - 55 and above
- To which gender identity do you most identify?
    - Male
    - Female
    - Prefer not to say
    - Prefer to self describe ______
- What is your academic discipline (e.g., Computer Science, Psychology, Medicine, etc.)? ____________
- What is the current status of your vision?
    - Normal or corrected to normal
    - Other ________________________ (briefly describe)

Interview questions

- How often do you commute using a public transit system such as a bus or train?
- What is the duration of your most frequent daily commute?
- What do you do during this commute time in a bus or train?
- How often do you read scholarly articles published in conferences or journals in your daily life?
- In general, what is your approach or strategy of reading such scholarly articles when you see it for the first time?
- What is the usual setting for you (time, place, comfort, etc.) to accomplish a scholarly article reading task?
- How often do you accomplish this task in mobile devices (smartphone or tablet)?
- If your strategy of reading a scholarly article is different from that of reading in a desktop or printed paper format, what are those strategies and why are they different?
- Can you think of a time when you had to read any scholarly article while commuting?
- Can you elaborate on your strategies of reading a scholarly article while commuting? What is it like for you to read while commuting?
- If you don't prefer reading while commuting, what are the reasons?
- What are the challenges of reading while commuting?
- What could better support these challenges for reading while commuting?

Observation probing

Following are the questions based on observing participants during a reading task.

- Can you explain your experience of listening to scholarly articles published in conferences or journals by elaborating on what worked well and what did not work well?
- How easy or difficult was it to listen and comprehend this (task) article using this (software used for the task) application?
- What would be the challenges of completing this task in a standing posture?
- What would be the challenges of completing this task in a sitting posture?
- I noticed that you did <an interaction technique> while reading <moment or instance in the material>. Could you please explain what was happening at that moment?

Default Question Block

Please tick the layout type of this task

Plain text	Original layout
O	O

Please tick the article type of this task

Research paper	Pew report
O	O

Block 1

Please state to what extent you agree or disagree with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I clearly understood the problem that this article addressed	O	O	O	O	O
I clearly understood the key takeaway message of this article	O	O	O	O	O
I clearly understood how this article is organised	O	O	O	O	O

Block 2

Please state to what extent you agree or disagree with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I clearly understood the topic of this article	O	O	O	O	O
I clearly understood the key takeaway or conclusion of this article	O	O	O	O	O
I clearly understood how this article is organised	O	O	O	O	O

Private Policy Terms of Use

Powered by Qualtrics

Task Questionnaire - Part 1

Participant: Session:

Click on each scale at the point that best indicates your experience of the task

Mental Demand

Low High

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

Low High

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

Low High

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance

Good **Poor**

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort

Low High

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration

Low High

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Continue >>

Appendix B

Study Materials Used in Preliminary Evaluation Study

In the following, we provide materials from our second study, preliminary evaluation of Skimmer in Chapter 6. Specifically, we provide the following -

- Call For Participation
- Consent Form
- Study Protocol
- Questionnaires

B.1 Call For Participation

Department of Computer Science
University of British Columbia
Vancouver, BC, V6T 1Z4

Are you a grad student? Are you interested in trying auditory skim reading on the bus?

Who are we?

We are researchers from the University of British Columbia: Dr. Joanna McGrenere, Dr. Dongwook Yoon, and Taslim Arefin Khan. We are conducting a study to understand the usefulness of auditory skim reading using a smartphone application.

We are looking for participants who ...

- are graduate students
- can understand and speak English fluently
- are familiar with reading scholarly articles published in conferences or journals in their respective research fields
- are 18 years or older

What is involved?

You will be using a smartphone application in a bus commuting setting. You will meet us at the UBC Nest. Then we will walk towards the UBC bus loop where we will board bus #99. Once the study tasks are completed, we will get off at the next stop and you will sit for an interview. The study will take approximately 60 minutes to complete, but we encourage participants to hold 90 minutes of their time.

Receive a $20 honorarium for this study

Interested in Participating?

If you are interested in participating or would like more information, please contact Taslim at [redacted].

Call for Participation (Version 2.1 / June 26, 2019)

B.2 Consent Form for Participants

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

Consent Form

Eyes-reduced document skimming on the go

Principal Investigator:
Joanna McGrenere, Professor, Department of Computer Science, University of British Columbia, [redacted].

Co-Principal Investigator:
Dongwook Yoon, Assistant Professor, Department of Computer Science, University of British Columbia, [redacted].
Taslim Arefin Khan, MSc student, Department of Computer Science, University of British Columbia, [redacted].

What the study is about:
The overall purpose of this research is to assess the usefulness of a text-to-speech skimming interface for eyes-reduced interaction in a bus commuting context.

What you will be asked to do: After you have read this document, I will respond to any questions or concerns that you may have. Once you have signed this consent form, you will be asked to:

- board on the bus #99 from UBC bus loop and find a seat
- interact with a smartphone and listen to an audio-based reading material
- get off the bus
- answer interview questions

This should take approximately 60 minutes (maximum 90 minutes) and be completed in 1 session.

The session may also be photographed, video, and/or audio recorded. You have the option not to be photographed/video/audio recorded.

Inclusion criteria: We are looking for participants who:

- are graduate students
- can understand and speak English fluently

- are familiar with reading scholarly articles published in conferences or journals in their respective research fields
- are 18 years or older.

Risks and Benefits: There are no known psychological, physical, and social risks involved in the proposed research. Since visually reading in a commuting context induces motion sickness in some people, participants may feel nauseous. Participants are welcome to pause/stop the study at any moment they feel nauseous. Participants may learn how to benefit from eyes-reduced skim reading while commuting to accomplish a reading task.

How the data collected will be used: Data collected (including any image, audio/video recordings) will be used for analysis. This analysis will help us understand our design and help us improve the design for future. The research findings from this data may also be used to publish scholarly articles in conferences or journals.

Compensation: You will receive $20 for this study. The compensation will be provided prior to the study as cash only.

Confidentiality: The results of your participation will be reported without any reference to you specifically. All information that you provide will be stored in Canada. Only the researchers involved in this research will have access to the data.

Your data may be used but will remain anonymous in any reports, research papers, thesis documents, and presentations that result from this work. Your name will never be used, only a code associated with the data. Any identifiable element in the images, videos, or audios will be blurred or obscured to prevent identification. The only link between the code associated and your actual name will be this consent form and a printed spreadsheet, that will be stored in a locked cabinet, in a research lab with controlled access at UBC. The file linking your name with the associated code will be kept separate from the data.

Data Retention: All electronic files will be stored in an encrypted hard drive of a password protected laptop. Backups of the data will be stored on the UBC Computer Science secure department servers. Any handwritten notes and paper transcripts will be kept in a locked cabinet in the researchers' laboratory or the Principal Investigator's office with controlled access at UBC. Five years after the completion of the research, the electronic data will be deleted, and all physical media, audio/video records, and paper transcripts will be shredded by the Principal Investigator.

Taking part is voluntary: Your participation is voluntary. You are free to stop your participation at any point. Refusal to participate or withdrawal of your consent or discontinued participation in the study will not result in any penalty or loss of benefits or rights to which you might otherwise be entitled. The Principal Investigator may at his/her discretion remove you from the study for a number of reasons. In such an event, you will not suffer any penalty or loss of benefits or rights which you might otherwise be entitled. It will not affect your current or future relationship with the University of British Columbia.

Ethical Concerns and Participant Rights

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Please tick at least one of the following

- ❏ I consent to be audio recorded in this study
- ❏ I consent to be video recorded in this study
- ❏ I consent to be photographed in this study
- ❏ I do not consent to be photographed, audio, and video recorded in this study

I, ________________________________, have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in this study. However, I realize that my participation is voluntary and that I am free to withdraw at any time.

____

Participant's Signature Date

B.3 Study Protocol and Participant Training

How you are gonna do it:

1. We will board a bus.
2. You will do one reading with Skimmer within a fixed amount of time. Try not to minimize your attention to the screen, i.e., look at the screen as less as possible.
3. We will unboard the bus.
4. Answer some questions.
5. We will board another bus.
6. You will do the second and final reading with VoiceDreamReader within a fixed amount of time. Try not to minimize your attention to the screen, i.e., look at the screen as less as possible.
7. We will unboard the bus.
8. Answer some questions.

What you should find: We want you to skim in a way so that you have answers to the following:

1. Gist of the article
2. Structural organisation of the article
3. What Research Questions were asked in the article
4. Main contributions of the article

Training: We trained participants by showing live demos of both Skimmer and VoiceDreamReader.

B.4 Questionnaires and Interview Questions

Observations on Key design elements in support of different sub-tasks

1. Sub-task: Listenability and listening experience
    a. Use of two voices
    b. Use of earcons to reinforce action-complete
    c. Embedding paralingual qualities
        i. Emphasizing important words/group of words
        ii. Emphasizing quotes
        iii. Itemized li5sts narration (pre-announcing number of items of follow for > 4 items)
        iv. Narrating statistical test results in the text
2. Sub-task: Navigation
    a. Sentence navigation
        i. Discreet jump
        ii. Continuous (or random)
    b. Discourse marker navigation
        i. Ambient earcon
    c. Paragraph navigation
    d. Section navigation
        i. Selection of a Section or Subsection
        ii. Transition (in/out) from the Overview page
    e. Spatial random access (start from anywhere)
    f. Speech rate variation
3. Sub-task: Eyes-free and momentary visual attention
    a. Single reading arrow head position in the middle
    b. Ephemeral display of numerical entity
    c. Current position notification
    d. Attention to visual entity
        i. Vibration nudge to catch attention
        ii. Spatial positioning of opt in/out button

Interview Questions

Open-ended questions:

1. What is your initial reaction or impression to using the app for audio skimming? In what ways did the app support / not support you?How did the app support skimming?

2. Is this app something that you can imagine yourself using in your life?

Example questions following participant score in the questionnaires

1. Can you please elaborate the reasoning behind your rating here?

2. It looks like you rated Physical Demand as the factor that contributed more to overall workload when using this application.. Why do you think so?

Example questions following observation on key design elements

1. How was your experience going to Conclusion from Introduction? How easy or difficult was it for you?

2. How was your experience of listening to statistical values? What kind of challenges did you face while listening to the numbers?

3. How was your experience of seeing figures and tables in the paper?

4. How easy or difficult was sentence/paragraph/discourse navigation? What challenges did you face during this navigation?

5. How did the two different voices (male and female) worked or not for your skimming tasks?

Task Questionnaire - Part 1

Participant: Session:

Click on each scale at the point that best indicates your experience of the task

Mental Demand

Low High

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

Low High

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

Low High

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance

Good **Poor**

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort

Low High

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration

Low High

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Continue >>